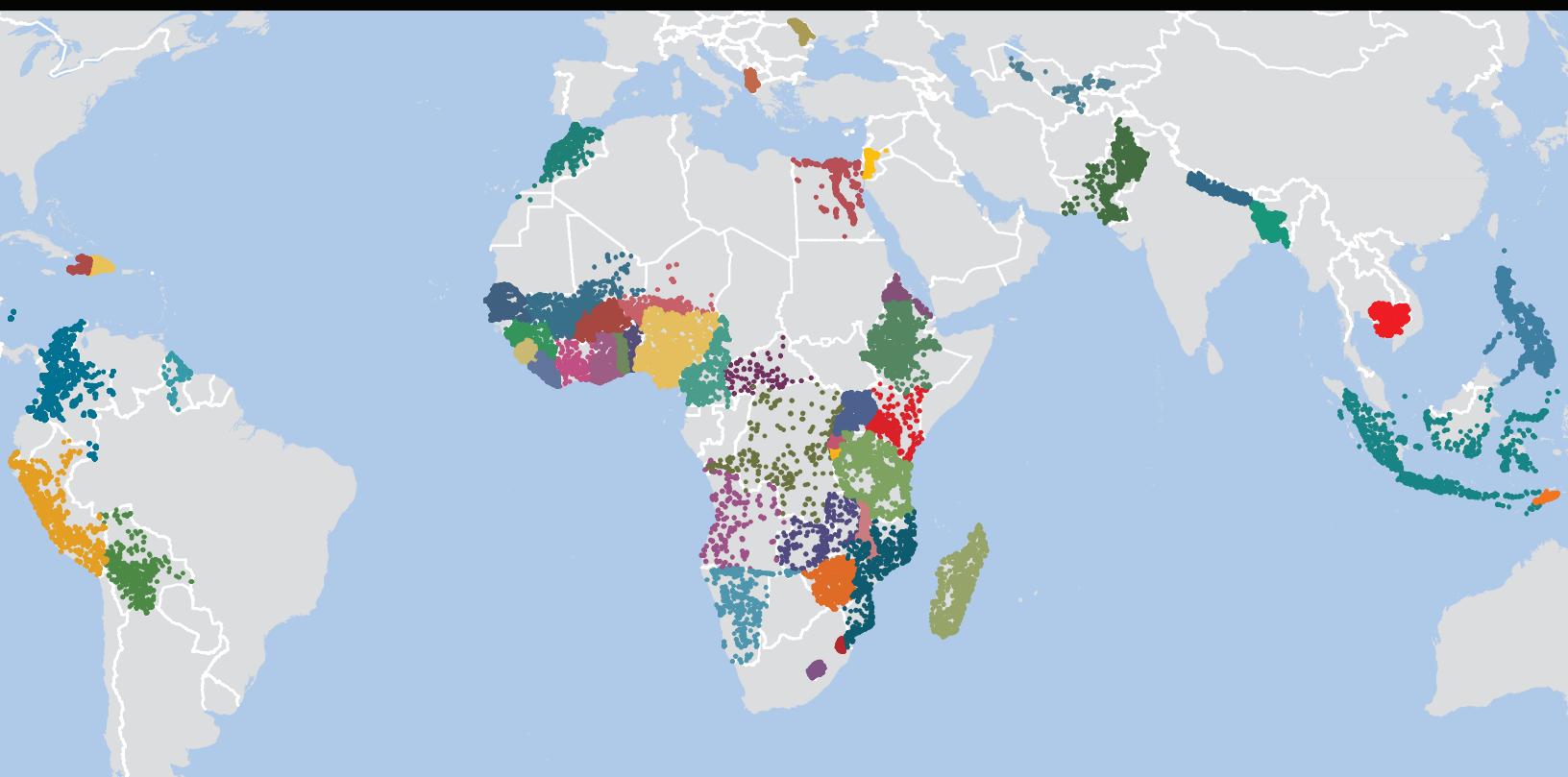




**USAID**  
FROM THE AMERICAN PEOPLE

# GUIDE D'UTILISATION DES CARTES DE SURFACES MODÉLISÉES DU DHS PROGRAM

## RAPPORTS D'ANALYSE SPATIALE DHS N°14



AOÛT 2016

Cette publication a été produite pour United States Agency for International Development. Elle a été préparée par Clara R. Burgert-Brucker, Trinadh Dontamsetti, Aileen M. J. Marshall, et Peter W. Gething.



# Rapports d'Analyse Spatiale DHS n°14

## **Guide d'utilisation des cartes de surfaces modélisées du DHS Program**

Clara R. Burgert-Brucker

Trinadh Dontamsetti

Aileen M. J. Marshall

Peter W. Gething

ICF International  
Rockville, Maryland, USA

Août 2016

*Auteur correspondant:* Clara R. Burgert-Brucker, International Health and Development, ICF International, 530 Gaither Road, Suite 500, Rockville, MD 20850, USA; phone: +1-301-572-0446; fax: +1-301-407-6501; email: clara.burgert@icfi.com

**Remerciements:** Les auteurs sont reconnaissants à Tom Pullum, Trevor Croft et Erica Nybro pour leur assistance et leurs orientations dans la préparation de ce rapport, ainsi qu'à Anne Linn, Maria Muniz et Rocco Panciera pour leur aide à la révision de ce rapport.

**Traduction:** Les auteurs voudraient remercier Abdramane Soura et Rosalind Howes pour la traduction du rapport en français et Monique Barrere pour l'édition du texte français.

Editeur: Diane Stoy

Production du Document: Natalie La Roche

Cette étude a été réalisée avec l'appui de l'Agence américaine pour le développement international (USAID) par le biais du programme DHS (#AID-OAA-C-13-00095). Les opinions exprimées sont celles des auteurs et ne reflètent pas nécessairement les points de vue de l'USAID ou du Gouvernement des Etats-Unis.

The DHS Program aide les pays à travers le monde dans la collecte et l'utilisation des données pour surveiller et évaluer des programmes de population, santé et nutrition. Pour plus d'informations sur The DHS Program, contacter: The DHS Program, ICF International, 530 Gaither Road, Suite 500, Rockville, MD 20850, USA. Téléphone: 301-407-6500; Fax: 301-407-6501; Email: [reports@dhsprogram.com](mailto:reports@dhsprogram.com); Internet: <http://www.dhsprogram.com>.

Citation recommandée :

Clara R. Burgert-Brucker, Trinadh Dontamsetti, Aileen M. J. Marshall, and Peter W. Gething. 2016. *Guide d'utilisation des cartes de surfaces modélisées du DHS Program*. Rapports d'Analyse Spatiale DHS n° 14. Rockville, Maryland, USA: ICF International.

## Table des matières

<b>Tables .....</b>	<b>v</b>
<b>Figures.....</b>	<b>v</b>
<b>Abréviations.....</b>	<b>vii</b>
<b>Glossaire .....</b>	<b>ix</b>
<b>Préface.....</b>	<b>xi</b>
<b>Résumé .....</b>	<b>xiii</b>
<b>Résumé long.....</b>	<b>xv</b>
<b>Introduction au document d'orientation.....</b>	<b>1</b>
<b>1 Pourquoi The DHS Program crée-t-il des surfaces modélisées?.....</b>	<b>3</b>
1.1 Contexte .....	5
1.2 Les fondements de l'interpolation spatiale .....	5
1.3 Interpolation spatiale et enquêtes auprès des ménages .....	6
1.4 Travaux antérieurs pertinents du DHS Program .....	7
<b>2 Quelles surfaces modélisées The DHS Program crée-t-il ? .....</b>	<b>9</b>
2.1 Surfaces modélisées standardisées.....	10
2.2 Disponibilité des surfaces modélisées.....	14
<b>3 Comment les surfaces modélisées du DHS Program sont-elles créées? .....</b>	<b>19</b>
3.1 Explication du processus de création de surfaces modélisées.....	20
3.2 Exemple de documentation de surface modélisée .....	26
3.3 Incertitude du modèle de surface et interprétation .....	30
<b>4 Quelles sont les limites et les hypothèses de la modélisation des surfaces ? .....</b>	<b>33</b>
4.1 Les limites du modèle .....	34
4.2 Différences de modélisation entre indicateurs et entre pays.....	36
4.3 Agrégation de l'estimation ponctuelle de la surface interpolée au niveau national ou infra-national du DHS Program.....	37
<b>5 Comment les surfaces modélisées du DHS Program peuvent-elles être utilisées ? .....</b>	<b>39</b>
5.1 Prise de décision avec des surfaces modélisées .....	40
5.2 Transformation de la carte de surface modélisée en produits utiles (opérationnalisation) .....	40
5.3 Considérations d'utilisation et limites .....	45
<b>Dernières réflexions .....</b>	<b>47</b>
Références.....	49



## Tables

Tableau 1.	Résumé des indicateurs inclus dans la diffusion des données de septembre 2016 .....	11
Tableau 2.	Résumé des covariables utilisées dans la modélisation des données publiées en septembre 2016 .....	14
Tableau 3.	Convention de nomination des surfaces modelisées.....	16
Tableau 4.	Exemple de paramètres du modèle, de contribution des covariables et de statistiques de validation .....	23
Tableau 5.	Estimation issue de l'ensemble des données par rapport à l'estimation issue du modèle de surface pour le retard de croissance chez les enfants en Tanzanie, par région de la DHS .....	38
Tableau 6.	Approches de suivi et d'évaluation des programmes passés avec des surfaces modélisées ....	41

## Figures

Figure 1.	Processus d'interpolation.....	6
Figure 2.	Résumé des statistiques de validation pour l'activité pilote .....	8
Figure 3.	Capture d'écran du site de téléchargement de surfaces modélisées.....	15
Figure 4.	Diagramme schématique résumant le processus de création de surfaces modélisées des données DHS .....	21
Figure 5.	Exemple de statistiques de données.....	22
Figure 6.	Exemple de diagramme de dispersion de validation .....	24
Figure 7.	Exemple de surface de l'estimation ponctuelle et de surface de l'incertitude du modèle .....	26
Figure 8.	Exemples de visualisation de surface d'incertitude du modèle .....	31
Figure 9.	Exemple d'estimation ponctuelle de surface en milieu urbain et surface de l'incertitude du modèle .....	35
Figure 10.	Diagrammes de dispersion à but de validation et statistiques de validation du retard de croissance chez les enfants .....	37
Figure 11.	Exemple d'agrégation de surface modélisée à l'échelle d'unités administratives .....	43
Figure 12.	Illustration du processus de calcul de l'estimation de la charge .....	44
Figure 13.	Illustration de l'extraction de la valeur de surface modélisée à la localisation des villages ....	45



## Abréviations

API	Application Program Interface
CPN	Consultation Prénatale
DHS	Demographic and Health Surveys
EAM	Erreur Absolue Moyenne
EDS	Enquêtes Démographique et de Santé
EQM	Erreur Quadratique Moyenne
GPS	Système de positionnement global
km	Kilomètre
MII	Moustiquaire Imprégnée d'Insecticide
MBG	Model-Based Geostatistics
ODD	Objectif de Développement Durable
PR2	R-carré Prédictif
RAS 9	Rapports d'Analyse Spatiale EDS n° 9 “Interpolation Spatiale avec les Données des Enquêtes Démographique et de Santé: Considérations Clés,” qui décrit les considerations clés pour la création des surfaces interpolées utilisant les données des EDS (Burgert 2014)
RAS 11	Rapports d'Analyse Spatiale EDS n° 11 “Création des Surfaces d'Interpolation Spatiale avec les Données des EDS”, qui décrit l'étude pilote conduite pour déterminer la pertinence de l'application des méthodes MBG pour interpoler les données des EDS (Gething et al. 2015)
UPE	Unité Primaire d'Echantillonnage
ZD	Zone de Dénombrement



## Glossaire

**Géo-masquage GPS de la DHS (déplacement):** Les grappes urbaines sont déplacées jusqu'à une distance de 2 kilomètres (km). Les grappes rurales sont déplacées jusqu'à 5 km, et 1 % des grappes rurales sélectionnées au hasard sont déplacées jusqu'à dix km. Des détails sur le processus de déplacement de données DHS géoréférencées et la variabilité spatiale des données résultantes se trouvent dans Burgert et al. 2013.

**Interpolation:** Méthode de création de nouvelles valeurs de données dans la plage des points de données connus.

**Modélisation géostatistique (MBG):** Une classe de modèles statistiques spatiaux pour l'interpolation de données ponctuelles géo-localisées. Les modèles MBG sont des modèles linéaires généralisés mixtes qui étendent la flexibilité des modèles de régression linéaire généralisée classiques (qui permettent d'adapter différents types de données non gaussiennes tels que les données de comptage ou de proportion, via une fonction de liaison, dans un espace gaussien) en utilisant une distribution normale multivariée pour représenter la variation spatiale ou spatio-temporelle.

**Surfaces modélisées:** Sortie spécifique des méthodes MBG utilisées pour la création des cartes spatiales interpolées avec des données EDS.

**Interpolation spatiale (carte d'interpolation spatiale):** Dans le cadre de ce document, l'interpolation spatiale fait référence à la technique générale ou au concept d'interpolation, mais ne fait pas référence à la méthode spécifique que nous utilisons, celle de la modélisation géostatistique (MBG).



## Préface

The Demographic and Health Survey (DHS) Program est l'une des principales sources de données internationales sur la fécondité, la planification familiale, la santé maternelle et infantile, la nutrition, la mortalité, la santé environnementale, le VIH/sida, le paludisme et l'offre de services de santé.

Les Rapports d'Analyse Spatiale des données des Enquêtes DHS complètent les autres séries de rapports DHS en répondant à l'intérêt croissant pour une perspective spatiale des données démographiques et de santé. Les principaux objectifs de toutes les séries de rapports DHS sont de fournir des informations pour la formulation des politiques au niveau international et d'évaluer les résultats de chaque pays dans un contexte international.

Les sujets dans les Rapports d'Analyse Spatiale DHS sont choisis par The DHS Program en consultation avec l'Agence américaine pour le développement international. Une gamme de méthodologies est utilisée, y compris des techniques géostatistiques et la statistique multivariée.

On espère que la série de Rapports d'Analyse Spatiale DHS sera utile aux chercheurs, aux décideurs de politiques et aux spécialistes des enquêtes, en particulier ceux qui travaillent dans les pays à revenu faible ou intermédiaire, et servira à améliorer la qualité et l'analyse des données d'enquête.

Sunita Kishor

Directeur, The DHS Program



## Résumé

Une meilleure compréhension de la variation géographique et des inégalités en matière de santé, de richesse et d'accès aux ressources à l'intérieur des pays est essentielle pour la réalisation des objectifs de développement durable. La surface modélisée du DHS Program contribue au besoin accru de la communauté de développement en matière d'estimations démographiques et de santé à petites échelles. The DHS Program met à la disposition du public un ensemble standard de surfaces spatialement modélisées pour chaque enquête en population, avec une liste sélective d'indicateurs pertinents pour la prise de décision en matière de santé, de démographie et de développement. Les surfaces modélisées sont créées à base d'informations géocodées sur les grappes des enquêtes démographiques et de santé actuelles et futures, ainsi qu'une sélection d'enquêtes antérieures. Les cartes sont publiquement disponibles et téléchargeables à partir du répertoire de données spatiales du DHS Program (<http://spatialdata.dhsprogram.com/>). Ce document d'orientation permettra aux utilisateurs de mieux comprendre les surfaces modélisées du DHS Program et leur utilisation potentielle dans la prise de décision. The DHS Program a adopté l'approche de la modélisation géostatistique ('Model-Based Geostatistics (MBG)' dans son acronyme anglais) pour créer les surfaces modélisées. C'est une méthode de création des surfaces par interpolation statistique rigoureuse qui génère de nouvelles valeurs de données pour des zones non échantillonnées à partir de points échantillonnés. Un nombre aussi important de surfaces modélisées pour un groupe varié d'indicateurs démographiques et de santé n'a jamais été réalisé par le passé et, à ce titre, les utilisations potentielles sont encore naissantes. De nombreux utilisateurs trouveront de nouvelles façons innovantes d'utiliser les surfaces modélisées qui ne sont pas décrites ou entièrement analysées dans ce document.



## Résumé long

Une meilleure compréhension de la variation géographique et des inégalités en matière de santé, de richesse et d'accès aux ressources à l'intérieur des pays est essentielle pour la réalisation des objectifs de développement durable. La surface modélisée du DHS Program contribue au besoin accru de la communauté de développement en matière d'estimations démographiques et de santé à petites échelles. The DHS Program met à la disposition du public un ensemble standard de surfaces spatialement modélisées pour chaque enquête démographique et de la santé, avec une liste sélective d'indicateurs pertinents pour la prise de décision en matière de santé, de démographie et de développement. Les surfaces modélisées sont créées à partir d'une base d'informations géocodées sur les grappes des enquêtes démographiques et de santé actuelles et futures, ainsi qu'une sélection d'enquêtes antérieures. Les cartes sont publiquement disponibles et téléchargeables à partir du répertoire de données spatiales du DHS Program (<http://spatialdata.dhsprogram.com/>).

Ce document d'orientation permettra aux utilisateurs de mieux comprendre les surfaces modélisées du DHS Program ainsi que leur utilisation potentielle dans la prise de décision. Ce document n'est pas une revue complète du processus de modélisation, ce qui est discuté dans d'autres documents; il ne fournit pas non plus une liste complète des utilisations potentielles des surfaces modélisées. Ce document a été rédigé à la fois pour les spécialistes en analyse géospatiale et les spécialistes en données non-géospatiales. Les spécialistes en analyse géospatiale trouveront des informations clés sur la création des surfaces modélisées, les limites des surfaces modélisées et comment opérationnaliser les surfaces modélisées pour diverses analyses géospatiales. Pour les non-spécialistes en analyse géospatiale, il existe des informations de base sur les surfaces modélisées et leur application dans l'analyse de leurs données.

L'approche adoptée par The DHS Program est la modélisation géostatistique ('Model-Based Geostatistics (MBG)' dans son acronyme anglais), une méthode de création de surfaces par interpolation statistique rigoureuse, qui génère de nouvelles valeurs de données en zones non échantillonnées, informée par les données des zones échantillonnées. Cette activité s'appuie sur plusieurs années de travail du DHS Program qui ont permis d'identifier les opportunités et les limites de la création de surfaces interpolées basées sur les données des enquêtes DHS, ainsi que sur une activité pilote qui a utilisé la modélisation géospatiale pour créer des surfaces pour trois pays et quatre indicateurs. Les surfaces modélisées sont produites à partir de données géoréférencées accessibles au public à partir du DHS Program ou d'autres sources de données spatiales pertinentes telles que les rasters environnementaux. Cela facilitera la réplication et la comparabilité entre les pays, ce qui favorisera la prise de décisions éclairées en matière politiques et programmatiques. Le résultat du modèle est une surface modélisée, avec une résolution de pixel de  $5 \times 5$  km. En complément, il existe des cartes de surface correspondantes qui estiment l'incertitude ou l'erreur potentielle associée aux surfaces modélisées.

Comprendre les hypothèses et les limites de la modélisation des surfaces est essentiel pour un usage et une interprétation correcte. Il existe plusieurs limites liées aux zones urbaines, à la temporalité et au biais de localisation. En outre, il existe des différences considérables dans la validité des surfaces modélisées entre pays et entre indicateurs à l'intérieur d'un même pays. Cela peut être dû à la distribution des zones échantillonnées, à la valeur agrégée de l'indicateur à l'échelle de la grappe et à la mesure dans laquelle les covariables sont les moteurs du processus mesuré dans chaque pays. Enfin, le processus de modélisation n'a pas spécifiquement ajusté le modèle pour recréer les estimations aux mêmes niveaux régionaux ou nationaux présentés dans les rapports finaux du DHS Program.

Les surfaces modélisées du DHS Program peuvent être utilisées pour suivre et évaluer des situations et des programmes et peuvent contribuer à la prise de décisions éclairées politiques et programmatiques futures. Ce qui est inclus dans ce document est une discussion des approches possibles pour opérationnaliser les surfaces modélisées telles que l'agrégation, l'estimation de la charge et la création de liens avec d'autres données. De nombreux utilisateurs trouveront de nouvelles façons innovantes d'utiliser les surfaces modélisées qui ne sont pas décrites ou entièrement analysées dans ce document. Un nombre aussi important de surfaces modélisées pour un groupe varié d'indicateurs démographiques et de santé n'a jamais été réalisé auparavant et, à ce titre, les utilisations potentielles sont encore naissantes. The DHS Program est impatient d'apprendre dans les années à venir comment d'autres personnes utiliseront les surfaces modélisées.

Les surfaces spatialement modélisées, créées par The DHS Program peuvent aider à répondre aux besoins des communautés nationales et internationales pour des estimations plus granulaires et spatialement détaillées que celles actuellement fournies par The DHS Program et la plupart des autres sources de données au niveau national. Ces types de cartes, que ce soit à l'échelle de la grille de  $5 \times 5$  km ou agrégées par la suite à l'échelle d'unités décisionnelles infra-nationales appropriées, peuvent fournir des informations nécessaires pour mesurer la variation géographique des indicateurs démographiques, de santé et de développement. Les surfaces spatialement modélisées du DHS Program offrent des informations supplémentaires qui aideront les décideurs à mieux comprendre la désagrégation géographique des principaux indicateurs démographiques et de santé dans les années à venir. Il existe un énorme potentiel pour des utilisations nouvelles et innovantes des surfaces modélisées. Ce n'est que dans une grande communauté d'utilisateurs qui partagent leurs expériences que ce potentiel sera pleinement réalisé.

## Introduction au document d'orientation

Depuis septembre 2016, The Demographic and Health Survey (DHS) Program a commencé à fournir un ensemble standard de surfaces spatialement modélisées pour une série d'indicateurs afin d'accompagner les enquêtes DHS actuelles et futures ainsi qu'une sélection d'enquêtes antérieures avec des informations sur grappes géo-codées. Les cartes sont publiquement disponibles et téléchargeables à partir du répertoire de données spatiales du DHS Program (<http://spatialdata.dhsprogram.com>). Les cartes sont produites à partir d'une combinaison de données DHS publiquement disponibles et de données externes disponibles à l'échelle mondiale utilisées dans la modélisation comme covariables. La modélisation applique des méthodes standardisées pour promouvoir la comparabilité entre pays et pour faciliter la prise de décisions dans la construction de politiques et programmes. Bien que la création de ces surfaces cartographiques ne soit pas nouvelle, leur incorporation dans un processus décisionnel plus formel n'est pas encore courante. Pour les utilisateurs secondaires des surfaces modélisées, il existe peu ou aucune orientation leur permettant de comprendre les opportunités et les limites dans leur utilisation malgré une demande accrue pour les cartes de surfaces modélisées. De nombreux groupes ont créé des surfaces à des fins diverses bien que l'utilisation opérationnelle de ces surfaces n'ait pas encore été complètement explorée. L'absence de guide d'utilisation et d'une documentation non technique publiée contraste avec la demande croissante pour ce type de matériel et l'augmentation du nombre de surfaces modélisées disponibles dans les domaines de la santé et de la démographie.

Ce document d'orientation permettra aux utilisateurs de mieux comprendre les surfaces modélisées du DHS Program ainsi que l'utilisation potentielle de ces surfaces pour la prise de décision. Ce document n'est pas une revue exhaustive du processus de modélisation, qui est abordé dans d'autres documents (Gething et al, 2015), et ne fournit pas non plus une liste complète des utilisations potentielles des surfaces modélisées. Étant donné qu'un nombre aussi important de surfaces modélisées pour un groupe diversifié d'indicateurs démographiques et de santé n'a jamais été réalisé auparavant, les utilisations potentielles de ces surfaces sont naissantes. De nombreux utilisateurs trouveront de nouvelles façons innovantes d'appliquer les surfaces modélisées qui ne sont pas incluses ou explorées dans ce document. The DHS Program est impatient d'apprendre dans les années à venir comment seront utilisées les surfaces modélisées.

### Public visé par le document d'orientation

Le document est destiné à la fois aux spécialistes en analyse géospatiale et aux spécialistes en données non-géospatiales.

- Les spécialistes en analyse géospatiale trouveront des informations clés sur la création des surfaces modélisées, les limites éventuelles de ces surfaces et les approches par lesquelles ces surfaces peuvent être opérationnalisées dans diverses analyses géospatiales.
- Les non-spécialistes en analyse géospatiale trouveront des informations de base sur les surfaces modélisées et comment ils peuvent utiliser ces surfaces dans les activités d'analyse de données.

Le document permet aux utilisateurs de sélectionner les parties qui leur paraissent pertinentes ou qui présentent un intérêt sans lire tout le document. Il est possible que certains utilisateurs trouvent certaines parties du document assez techniques, en particulier la section 3 dont la lecture n'est pas essentielle pour tous les lecteurs. Il est important cependant que tous les utilisateurs comprennent les limites et les hypothèses, abordées dans la section 4, qui leur permettront d'utiliser les surfaces modélisées de manière appropriée.

## Organisation du document

Le document comprend cinq grandes sections dont chacune répond à une question principale:

1. Pourquoi The DHS Program crée-t-il des surfaces modélisées ?
2. Quelles sont les surfaces modélisées créées par The DHS Program ?
3. Comment les surfaces modélisées du DHS Program sont-elles créées ?
4. Quelles sont les limites et les hypothèses des surfaces modélisées ?
5. Comment les surfaces modélisées du DHS Program peuvent-elles être utilisées ?

Chaque section commence par un résumé des concepts abordés dans la section ainsi que des «questions clés». Ces questions mettent en évidence les problèmes importants dans chaque section, avec des réponses courtes qui sont abordées plus en profondeur dans la partie suivante.

## Qu'est-ce que The DHS Program ?

The Demographic and Health Survey (DHS) Program, est depuis longtemps un chef-de-file dans la collecte et la dissémination de données d'enquêtes aléatoires par grappes concernant les principaux indicateurs de développement (<http://dhsprogram.com/>). Outre les fichiers de données standards accessibles au public dans lesquels les résultats des enquêtes des individus et de ménages peuvent être compilés par région infra-nationale (province ou état) et par strate urbaine/rurale, la plupart des enquêtes fournissent désormais des données géo-codées pour chaque grappe échantillonnée (zones de dénombrement (ZD)). Les coordonnées du système de positionnement global (GPS) des grappes dans les enquêtes DHS auprès des ménages fournissent des informations à l'échelle locale qui peuvent être liées aux résultats de l'enquête pour quantifier les variations et les inégalités démographiques et sanitaires.

# 1 Pourquoi The DHS Program crée-t-il des surfaces modélisées?

## Résumé

Les surfaces modélisées du programme des Enquête DHS contribuent aux nombreux besoins de la communauté du développement en matière d'estimations démographiques et sanitaires sur petites échelles. Une meilleure compréhension de la variation géographique et de l'inégalité en matière d'état de santé, de richesse et d'accès aux ressources à l'intérieur des pays est essentielle pour atteindre les Objectifs de Développement Durable (ODD). The DHS Program a adopté l'approche de la modélisation géostatistique (MBG) pour créer des surfaces modélisées. L'approche MBG est une méthode de création de surfaces par interpolation statistique rigoureuse qui crée de nouvelles valeurs d'indicateurs pour les zones non échantillonnées à partir de données échantillonnées. Cette activité s'appuie sur plusieurs années de travail du DHS Program, axé sur l'identification des opportunités et des limites de la création de surfaces interpolées avec des données des enquêtes DHS, et la conduite d'une activité pilote utilisant l'approche MBG pour la création des surfaces modélisées pour trois pays et quatre indicateurs.

## Questions clés

### ***Qu'est-ce que l'interpolation?***

L'interpolation est une approche statistique dans laquelle des valeurs prédictives sont établies pour des localités non-échantillonnées sur la base d'une combinaison pondérée de points de données voisins. Voir la section 1.2.

### ***Ces surfaces remplaceraient-elles une grande enquête?***

Non, ces surfaces modélisées utilisent les données d'enquête du DHS Program. Sans ces données d'enquête, il n'y aurait pas de données disponibles pour créer les cartes. Voir la section 1.3.

### ***Ces méthodes de surface admettent-elles des tailles d'échantillon plus petites ?***

Ces approches peuvent admettre des tailles d'échantillon plus petites dans certains pays car de gros échantillons ne sont pas nécessaires pour estimer les résultats au niveau administratif. Les petits échantillons conduisent cependant à des surfaces plus imprécises. La décision d'utiliser des échantillons de plus petite taille dépend également des besoins du pays et des contraintes budgétaires. Voir la section 1.3.

### ***Pouvez-vous donner des informations sur les lieux d'insécurité et où nous ne pouvons pas aller faire l'enquête?***

Oui, ces techniques permettent des estimations dans des zones non-enquêtées. Ces zones ne seraient pas modélisées avec autant de certitude que les autres zones échantillonnées. Cependant, en fonction de leur taille et du type d'insécurité, ces zones peuvent avoir des résultats démographiques et de santé différents de ce qui est observé dans les zones étudiées. Voir la section 1.3.

### ***Pourquoi cela n'a-t-il pas déjà été fait par The DHS Program?***

Les techniques de modélisation utilisées pour ces surfaces et appliquées aux indicateurs démographiques et de santé sont relativement nouvelles et, jusqu'à récemment, elles ont été développées par des groupes universitaires comme preuve de concept plutôt que pour l'action politique et la prise de décision. De plus, ces techniques de modélisation s'appuient sur des covariables spatiales externes qui sont devenues plus courantes et accessibles au public au cours des dernières années. Les techniques ont également été rationalisées pour nécessiter moins de calcul dans un contexte où la puissance de calcul informatique est devenue plus grande et plus disponible. Voir la section 1.4.

## 1.1 Contexte

Une meilleure compréhension de la variation géographique et de l'inégalité en matière de santé, de richesse et d'accès aux ressources à l'intérieur des pays est de plus en plus reconnue comme essentielle pour atteindre les ODD. Les indicateurs de développement et de santé évalués au niveau national peuvent souvent dissimuler des inégalités importantes entre des zones administratives/ géographiques plus petites avec souvent des populations pauvres des zones rurales les moins bien représentées. Alors que le financement international pour la santé et le développement est de plus en plus sous pression, la capacité d'orienter les ressources limitées vers les groupes mal desservis devient plus cruciale. En même temps, il existe des retards dans les progrès pour la réalisation des objectifs pour les indicateurs clés de la santé mondiale. Le suivi des inégalités dans le domaine démographique, en terme d'état de santé et d'accès à la santé pour cibler les interventions et mesurer les progrès vers la réalisation des objectifs de santé et de développement tels que les ODD nécessite des preuves fiables, détaillées et désagrégées. En outre, à mesure que les gouvernements nationaux se décentralisent et que les décisions politiques sont prises au niveau local à de petites échelles administratives, il est de plus en plus nécessaire d'utiliser les données existantes pour cibler, suivre et évaluer précisément l'impact des programmes dans des zones géographiques plus petites. Actuellement, trois approches permettent d'estimer des indicateurs provenant d'enquêtes en population sur les petites unités géographiques :

1. Agrandissement du processus de collecte de données d'enquêtes nationales représentatives en augmentant la taille de l'échantillon, les coûts d'enquête et le temps d'enquête nécessaire pour créer un échantillon représentatif au niveau administratif souhaité.
2. Utilisation des données provenant des systèmes d'information sanitaire de routine des établissements de santé ou des communautés.
3. Estimation sur de petites échelles, comprenant des cartes spatialement interpolées qui utilisent la modélisation et des techniques statistiques pour prédire des valeurs pour de petites unités géographiques.

La première approche n'est pas souvent réalisable dans un environnement de plus en plus limité en ressources. La qualité des données de la deuxième option n'est pas toujours fiable. Les données ne sont pas facilement accessibles et ne sont généralement pas représentatives à l'échelle nationale. C'est la troisième approche utilisant l'interpolation spatiale qui a suscité un intérêt accru au cours des dernières années.

## 1.2 Les fondements de l'interpolation spatiale

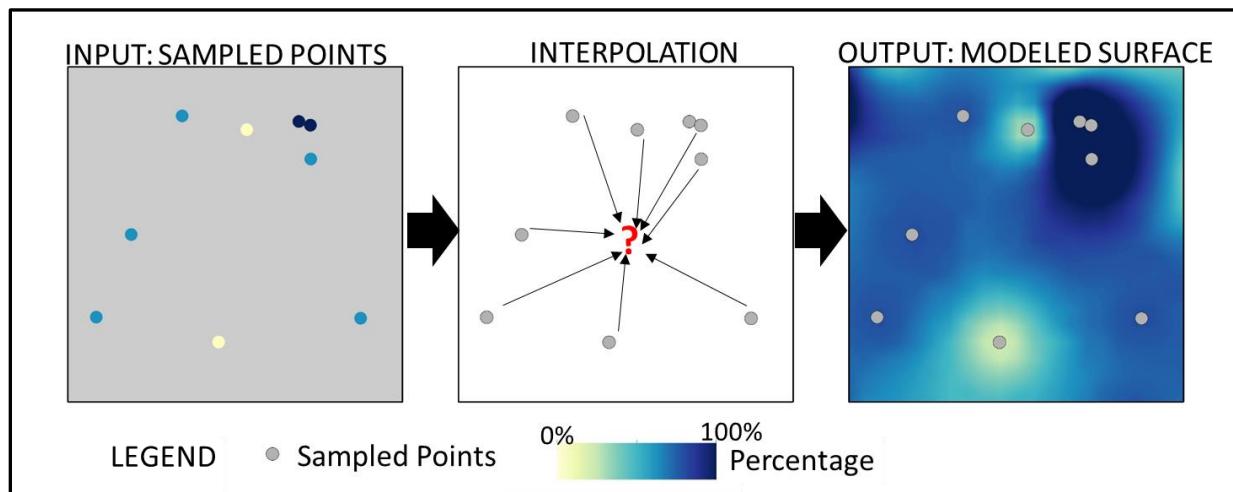
Le terme *géostatistique* fait référence à une collection d'outils statistiques qui aident à la compréhension et à la modélisation de la variabilité spatiale. La motivation principale est la prédiction des valeurs non-échantillonnées dispersées dans l'espace (interpolation) (figure 1). La méthode la plus souvent utilisée, *le krigage*, est une approche d'interpolation dans laquelle des valeurs prédites sont générées pour des localités non-échantillonnées sur la base d'une combinaison pondérée de points de données proches. Contrairement aux algorithmes d'interpolation plus simples, le krigage fournit une précision optimale des valeurs prédites en identifiant les poids les plus appropriés pour chaque point de données. Ceci est réalisé en caractérisant le degré de corrélation entre les points à travers l'espace avec un variogramme.

*L'inférence bayésienne* est une méthode d'inférence statistique basée sur le théorème de Bayes. Cela permet la combinaison de connaissances antérieures avec de nouvelles informations. L'inférence

bayésienne est largement utilisée comme une approche flexible, théoriquement rigoureuse, pour construire des modèles statistiques basés sur des ensembles de données échantillonnées.

La *géostatistique bayésienne* fait référence à la mise en œuvre de modèles géostatistiques avec des méthodes d'inférence bayésiennes. L'incertitude des données due à la variation d'échantillonnage et aux paramètres du modèle ajusté (tels que la forme du variogramme ou de la fonction d'autocorrélation et les relations avec les covariables) est inférée et propagée de sorte qu'elle peut être mesurée dans les predictions faites. En termes pratiques, cela fournit un moyen pratique de propager l'incertitude à travers toutes les étapes de l'ajustement du modèle et de représenter sur carte cette incertitude en tant que distribution postérieure pour chaque valeur de pixel prédicté.

**Figure 1. Processus d'interpolation**



### 1.3 Interpolation spatiale et enquêtes auprès des ménages

Les techniques d'interpolation spatiale pour estimer des valeurs sur de petites unités géographiques ne remplacent pas la nécessité d'enquêtes représentatives au niveau national. La principale variable d'entrée dans le processus de l'interpolation spatiale est la valeur de l'indicateur pour chaque localité géo-référencée dans la DHS. Ces techniques peuvent admettre des tailles d'échantillon plus petites dans certains pays étant donné que de gros échantillons ne sont pas nécessaires pour estimer des résultats sur de petites unités géographiques. Cependant, les estimations avec des échantillons de petite taille donnent des cartes de surface plus imprécises. La taille de l'échantillon elle-même dépend finalement de facteurs comprenant les types d'indicateurs mesurés (les événements rares comme la mortalité nécessitent une plus grande taille d'échantillon), le niveau de représentativité de l'échantillon (national versus sub-national), les autres besoins de l'enquête et les contraintes budgétaires.

Les techniques d'interpolation spatiale permettent d'estimer les valeurs des indicateurs dans les localités qui n'ont pas été enquêtées en général et dans les zones exclues de l'enquête en raison de l'insécurité. Cependant, les grandes zones non-échantillonnées ne seraient pas modélisées avec autant de précision que d'autres zones échantillonnées. En outre, selon leur taille et le type d'insécurité, ces zones peuvent avoir des résultats démographiques et de santé différents de ce qui est observé dans les zones enquêtées.

## 1.4 Travaux antérieurs pertinents du DHS Program

### 1.4.1 Rapports d'Analyse Spatiale EDS n° 9

The DHS Program a convoqué une réunion des acteurs clés en juin 2013 pour discuter de l'utilisation des données géographiques issues des DHS pour l'interpolation spatiale. Cette réunion s'est tenue dans le contexte suivant :

- Avancées technologiques comprenant une puissance de calcul plus rapide, des données GPS plus accessibles et un matériel informatique moins coûteux.
- Désir des décideurs d'utiliser des données sur de petites unités administratives et sur d'autres zones géographiques pertinentes (zones de moyens de subsistance).
- Utilisation et adoption accrues des données DHS dans les techniques de modélisation spatiale en général et plus particulièrement dans les techniques d'interpolation.

Les participants ont également discuté des particularités des données géoréférencées des DHS auprès des ménages, qui pourraient limiter certaines applications. Le Rapports d'Analyse Spatiale DHS n° 9 (RAS 9) (Burgert 2014) résume les principales discussions et recommandations de cette réunion qui s'est intéressée également à la sélection des indicateurs, aux méthodologies et aux limites.

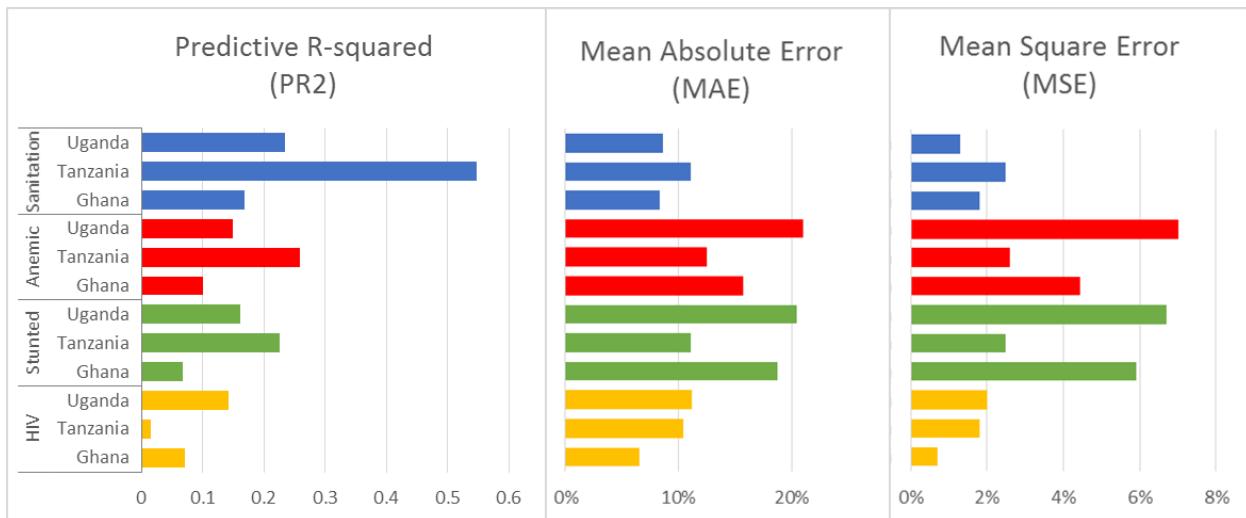
### 1.4.2 Rapports d'Analyse Spatiale EDS n° 11

Après la réunion de juin 2013, The DHS Program a commencé à explorer l'utilisation potentielle de la modélisation géostatistique bayésienne (MBG) pour la production de surfaces modélisées, interpolées à partir des données GPS des grappes des enquêtes EDS. Comme preuve de concept, les méthodes MBG ont été testées sur quatre indicateurs provenant de trois DHS : DHS 2008 du Ghana, DHS 2010 de la Tanzanie, et DHS 2011 de l'Ouganda. Les quatre indicateurs comprenaient la prévalence du dépistage du VIH chez les femmes (au cours des 12 derniers mois), les prévalences du retard de croissance et de l'anémie chez les enfants, et l'accès des ménages à un assainissement amélioré. Le Rapports d'Analyse Spatiale EDS n° 11 (RAS 11) (Gething et al., 2015) résume les résultats détaillés de cette activité de démonstration et inclut l'évaluation de la validité de la méthode, des covariables et de l'incertitude. En général, à l'exception de l'indicateur concernant le dépistage du VIH, les modèles ont raisonnablement bien fonctionné pour tous les indicateurs avec de faibles biais et des erreurs moyennes inférieures à 20 points de pourcentage, et plus proches de dix points dans la plupart des cas (Figure 2). Les faibles valeurs de l'erreur quadratique moyenne (EQM) indiquent que le modèle a un biais global minimal. Il n'est pas surprenant que les variations géographiques d'une variable telle que l'accès au dépistage du VIH qui a principalement des déterminants non-biophysiques soient moins réceptives à la capture de covariables principalement environnementales, avec généralement de faibles valeurs prédictives du R-carré (PR2). Malgré cela, les erreurs absolues moyennes (EAM) étaient relativement faibles.

En outre, le rapport explore l'impact du géo-masquage des grappes DHS géo-référencées sur la production de surfaces interpolées. Une résolution de pixel de  $5 \times 5$  km a été choisie pour réduire l'impact du géo-masquage des grappes DHS géo-référencées sur la surface finale fournie par le modèle et pour correspondre à la résolution des covariables d'entrée du modèle. Enfin, le rapport s'est intéressé au potentiel de nouvelles méthodologies et covariables pour relever le défi de la cartographie des zones urbaines. Le résultat du modèle complet de l'activité de démonstration du concept comprenait 12 surfaces

modélisées avec des pixels carrés de  $5 \times 5$  km pour la surface cartographiée prédite et la carte de surface de l'incertitude.

**Figure 2. Résumé des statistiques de validation pour l'activité pilote**



## 2 Quelles surfaces modélisées The DHS Program crée-t-il ?

### Résumé

The DHS Program met à la disposition du public un ensemble standard de surfaces spatialement modélisées pour chaque enquête de population pour une liste sélective d'indicateurs (Tableau 1). Chaque surface modélisée sera produite avec des méthodes de modélisation géostatistique standardisées (un type d'interpolation spatiale) et un ensemble standardisé de covariables à travers les pays (Tableau 2). Les surfaces modélisées sont produites à partir de données géoréférencées accessibles au public grâce au DHS Program et d'autres sources de données spatiales pertinentes (rasters environnementaux) (Tableau 2). Cela facilitera la réplication et la comparabilité entre les pays, ce qui favorisera la prise de décisions éclairées en matière de politiques et de programmes. Il y aura également des cartes de surfaces correspondantes aux estimations de l'incertitude ou de l'erreur potentielle associée aux surfaces modélisées. Les cartes sont publiquement disponibles pour le téléchargement sur le répertoire des données spatiales du DHS Program, à <http://spatialdata.dhsprogram.com/>.

### Questions clés

#### *Pourquoi n'utilisez-vous pas les données de recensement détaillées de mon pays comme données d'entrée du modèle ?*

The DHS Program a choisi d'utiliser une approche standardisée applicable dans différents pays et dans le temps. Cela signifie que des fichiers de données de covariables disponibles au niveau mondial sont utilisés au lieu de données spécifiques au pays telles que les données des recensements ou d'autres données de santé. Les fichiers de données par pays ne sont pas toujours disponibles publiquement et l'utilisation de fichiers de données standardisés au niveau mondial est plus facile lors de la production d'un grand nombre de fichiers de données au fil du temps. Voir la section 2.1.

#### *Pouvez-vous faire cela pour mon sujet spécifique ?*

Théoriquement, l'approche de modélisation peut être appliquée à un large éventail de sujets, bien qu'il y ait des limites pour certains types d'indicateurs qui pourraient ne pas être aussi bien modélisés avec ces techniques. Ces limites sont résumées dans cette section et expliquées plus en détail dans le RAS 9 (Burgert 2014). Voir la section 2.1.1.

#### *Pourquoi une seule enquête est-elle utilisée à la place de toutes les enquêtes menées dans un pays ?*

L'incorporation d'enquêtes réalisées à différents moments dans le temps nécessite l'utilisation d'une interpolation spatio-temporelle qui est considérablement plus complexe et nécessite souvent des covariables variant dans le temps qui ne sont pas toujours disponibles. En outre, la disponibilité de plusieurs enquêtes varie d'un pays à l'autre et leur incorporation aurait conduit à moins de standardisation entre les pays. Par souci de simplicité et de cohérence, une seule enquête par pays a été modélisée. Voir la section 2.1.2.

#### *Les covariables ont-elles la même date de référence que l'enquête ?*

Les covariables ont des dates de références différentes, bien qu'elles soient généralement proches de la date de l'enquête. Le tableau 2 résume les détails des covariables utilisées dans la modélisation. Voir la section 2.1.3.

#### *Les ensembles de données cartographiques sont-ils accessibles au public ? Où peut-on les télécharger ?*

Oui, les surfaces modélisées sont disponibles au téléchargement sur le répertoire de données spatiales du DHS Program à l'adresse <http://spatialdata.dhsprogram.com/>. Voir la section 2.2.

## 2.1 Surfaces modélisées standardisées

Il existe une distinction importante entre créer les «meilleures cartes possibles» pour un pays spécifique et créer des «cartes standardisées» pour le monde entier ou un ensemble de pays. La construction de la meilleure carte possible pour un pays donné pourrait nécessiter l'utilisation de nombreux ensembles de données contenant des covariables spécifiques au pays (telles que les données du recensement national) ainsi qu'une résolution et des méthodes de cartographie sur mesure. Bien qu'elles puissent potentiellement optimiser la précision des cartes, ces données spécifiques au pays empêcheraient inévitablement la comparaison directe avec les cartes d'autres pays ayant des données différentes. Seules les covariables disponibles pour l'ensemble des pays ont été incluses dans les modèles afin d'assurer une approche standardisée à travers tous les pays. En outre, The DHS Program utilise des données géoréférencées accessibles au public à la fois à partir des enquêtes DHS (emplacements GPS) et des sources de covariables externes de sorte que les surfaces puissent être reproductibles. En pratique, cela signifie que les covariables (Tableau 2) provenaient principalement de la télédétection par satellite, bien que ces données aient été largement utilisées et considérées comme prédicteurs utiles d'un large éventail d'indicateurs sociaux, économiques et de santé.

### 2.1.1 Sélection des indicateurs pour une série de cartes standard

Tous les indicateurs qui peuvent être dérivés d'un fichier de données DHS peuvent ne pas être appropriés pour être utilisés dans un processus de création de surface modélisée. Le Rapport d'Analyse Spatiale EDS n° 9 (RAS 9) (Burgert 2014) récapitule les caractéristiques suivantes pour la sélection d'indicateurs qui conviennent à tout type de technique d'interpolation spatiale:

- L'indicateur est bien mesuré dans les enquêtes DHS et est une mesure robuste qui n'est pas sujet à une erreur de mémoire significative.
- L'indicateur ne mesure pas un événement rare comme la mortalité néonatale. Ceci est important car la première étape du processus de modélisation de surface nécessite le calcul de l'indicateur au niveau de la grappe. Un événement relativement rare peut avoir plusieurs localités avec des données nulles. Bien que des valeurs de numérateur/dénominateur nulles ou généralement faibles soient acceptées dans le modèle, trop de cas de zéro dans un fichier de données conduiront à des surfaces avec une incertitude considérable qui limite l'utilisation de la surface.
- L'indicateur est spatialement hétérogène et varie selon l'espace géographique.
- L'indicateur correspond à une période de référence spécifique-pas une période de référence indéfinie ou une période de référence spatialement liée au résultat (jamais testé pour le VIH par rapport à la présence de la fièvre au cours des deux dernières semaines).
- L'indicateur n'est pas soumis à des restrictions temporelles ou micro-saisonnieres et, par conséquent, il est peu probable qu'il change sensiblement au cours de la collecte des données qui peut durer plusieurs mois. Par exemple, la fréquentation scolaire et l'utilisation de moustiquaires ont été exclues en raison de problèmes de temporalité. Cela signifie que les localités peuvent être enquêtées dans la même zone physique à différents moments. De plus, dans un pays tout entier, différents déterminants saisonniers peuvent compliquer la compréhension de certains indicateurs temporellement liés. Par exemple, des activités de campagne telles que la vaccination ou la distribution de moustiquaires peuvent également avoir lieu à différents moments dans différents endroits.

- L'indicateur se rapporte à la localisation actuelle de l'enquêtée et non à la mortalité maternelle par la méthode des soeurs qui repose sur l'interrogation des enquêtées sur la survie de toutes leurs soeurs adultes.

Sur la base de ces critères, un ensemble initial de 15 indicateurs a été sélectionné pour créer des surfaces de cartes spatiales en utilisant les données DHS, pour une diffusion publique en septembre 2016. Ces indicateurs sont résumés dans le Tableau 1. La définition standard de l'indicateur a été utilisée pour chaque indicateur, avec les estimations nationales comparées aux valeurs du programme API de la DHS (<http://api.dhsprogram.com/>) avant le début de la modélisation. Des informations supplémentaires sur les indicateurs individuels et la manière dont ils sont collectés sont disponibles sur le site Web du DHS Program (<http://dhsprogram.com/>). Ces indicateurs étaient pertinents pour l'ensemble de la communauté du développement, y compris les ODD et d'autres priorités programmatiques, et importants pour équilibrer les indicateurs au niveau du ménage et de l'individu (femmes, hommes et enfants). Cette liste peut changer au fil du temps, avec des indicateurs ajoutés ou supprimés à mesure que leur pertinence pour la communauté du développement en général sera évaluée et leur utilité globale dans la prise de décision potentielle mieux comprise. Tous les indicateurs ne sont pas disponibles pour chaque enquête, soit parce que les données appropriées ne sont pas disponibles (tous les pays n'effectuent pas l'enquête auprès des hommes), soit parce que l'indicateur est recueilli de manière non standard dans le pays, comme par exemple, les estimations sur la période de référence étant de 3 ans ou de 5 ans.

**Tableau 1. Résumé des indicateurs inclus dans la diffusion des données de septembre 2016**

Indicateur	Définition
<b>Population vivant dans un ménage utilisant une source d'eau améliorée</b>	Pourcentage de la population (de jure) vivant dans un ménage dont la principale source d'eau potable est une source améliorée
<b>Population vivant dans un ménage n'utilisant pas de toilettes (défécation dans la nature)</b>	Pourcentage de la population (de jure) vivant dans un ménage dont le principal type de toilettes est "aucune installation" (défécation à l'air libre).
<b>Personnes ayant accès à une MII</b>	Pourcentage de la population (de facto) du ménage qui pourrait dormir sous une MII si chaque MII du ménage était utilisée par deux personnes au maximum.
<b>Femmes en union utilisant actuellement une méthode de contraception moderne</b>	Pourcentage de femmes actuellement mariées ou en union utilisant actuellement une méthode moderne de contraception.
<b>Demande en planification familiale satisfait par les méthodes modernes</b>	Le pourcentage de la demande en planification familiale satisfait par les méthodes modernes est calculé comme le nombre de femmes en union utilisant les méthodes modernes de planification familiale divisé par le nombre de femmes actuellement en union avec une demande de planification familiale (soit avec des besoins non satisfaits ou utilisant une méthode de planification familiale).
<b>Besoin non satisfait en planification familiale</b>	Pourcentage de femmes actuellement mariées ou en union ayant des besoins non satisfaits en matière de planification familiale.
<b>Femmes 15-49 ans anémiques</b>	Pourcentage de femmes classifiées comme ayant de l'anémie (<12,0 g / dl pour les femmes non enceintes et <11,0 g / dl pour les femmes enceintes).
<b>Visites prématernelles pendant la grossesse: 4 visites ou +</b>	Pourcentage de femmes ayant eu une naissance vivante au cours des cinq années précédant l'enquête qui ont eu 4 visites prématernelles ou plus.
<b>Lieu d'accouchement: Établissement de santé</b>	Pourcentage de naissances vivantes des 5 années avant l'enquête qui se sont déroulées dans un établissement de santé.
<b>Femmes alphabétisées</b>	Pourcentage de femmes âgées de 15 à 49 ans alphabétisées.
<b>Vaccination DTC3 reçue</b>	Pourcentage d'enfants de 12 à 23 mois ayant reçu une troisième dose de DTC.
<b>Vaccination contre la rougeole reçue</b>	Pourcentage d'enfants de 12 à 23 mois qui ont été vaccinés contre la rougeole.

	rougeole.
<b>Enfants présentant un retard de croissance</b>	Pourcentage d'enfants de moins de cinq ans présentant un retard de croissance (inférieur à -2 ET de la taille pour âge selon les normes de l'OMS).
<b>Hommes alphabétisés</b>	Pourcentage d'hommes âgés de 15 à 49 ans alphabétisés.
<b>Utilisation du tabac parmi les hommes</b>	Pourcentage d'hommes âgés de 15 à 49 ans qui consomment du tabac.

### **2.1.2 Considérations relatives aux données des DHS**

Dans la plupart des enquêtes DHS auprès des ménages, les grappes d'échantillonnage constituent l'Unité Primaire d'Echantillonnage (UPE), qui comprend des zones géographiques préexistantes appelées Zones de Dénombrement (ZD). Les limites des ZD sont définies par le bureau de recensement du pays de même que le statut urbain et rural de chaque grappe. Une ZD peut être un carré de la ville (bloc de maisons) ou un immeuble d'appartements dans les zones urbaines tandis que dans les zones rurales, une ZD est généralement un village ou un groupe de villages. La population et la taille des grappes échantillonnées varient entre les pays et à l'intérieur de ceux-ci. En règle générale, les grappes contiennent 100 à 300 ménages, parmi lesquels 20 à 30 ménages sont sélectionnés au hasard pour participer à l'enquête. Le centre estimé de chaque grappe est enregistré comme une coordonnée en latitude/longitude, obtenue à partir d'un récepteur GPS ou dérivée de cartes publiques en ligne ou de répertoires toponymiques. La taille physique réelle ou les limites de la grappe sont publiquement disponibles bien qu'au cours des dernières années, il est devenu plus courant pour les pays d'avoir des fichiers de limites des ZD qui sont utilisés pour calculer le centre de la ZD.

Pour assurer la confidentialité, les emplacements de grappes géocodées sont géo-masqués (déplacés) avant la diffusion des fichiers de données (Burgert et al., 2013). Les grappes urbaines sont déplacées jusqu'à une distance de deux kilomètres. Les grappes rurales sont déplacées sur une distance pouvant aller jusqu'à cinq kilomètres, et 1 % des grappes rurales sélectionnées au hasard sont déplacées à une distance pouvant aller jusqu'à dix kilomètres. Le processus de création des surfaces modélisées utilise les fichiers de données géo-masqués rendus publics par The DHS Program.

L'approche de la modélisation a utilisé une seule enquête DHS comme données d'entrée. Il est possible de créer des modèles spatiotemporels qui utilisent plusieurs enquêtes comme données d'entrées pour créer une seule surface. Cette approche peut augmenter le pouvoir prédictif du modèle, mais nécessite un certain nombre d'autres considérations telles que la date des enquêtes et les dates de référence des covariables. Pour les besoins des surfaces standard spatialement modélisées des enquêtes DHS, une approche par année d'enquête a simplifié le traitement et l'interprétation du modèle.

### **2.1.3 Considérations sur les covariables géospatiales**

Un aspect important de la modélisation géostatistique est l'exploitation des covariables géospatiales qui sont pertinentes ou qui sont liées à l'indicateur d'intérêt, qui peuvent expliquer partiellement la variation de cet indicateur et permettre des prédictions plus précises à travers la carte. Comme indiqué ci-dessus, des données internationales sur les covariables ont été utilisées à la place des données spécifiques au pays, car cela garantit la normalisation entre les pays. Une série de covariables géospatiales a été choisie à partir de la littérature existante qui a déjà démontré leur grande utilité dans la cartographie géospatiale (Gething et al., 2015; Weiss et al., 2014). Les sources de données géospatiales décrites dans le Tableau 2 ont été obtenues dans une variété de résolutions spatiales et d'étendues géographiques. (Les détails sur les covariables sont disponibles sur le site Web pour chaque type de données inscrit dans le tableau.) De plus, les formats terre-mer différaient légèrement entre les produits, de sorte que la définition précise des côtes

et l'inclusion ou l'exclusion des petites îles et des péninsules n'étaient pas cohérentes. Ces facteurs ont empêché l'utilisation directe de ces données dans un modèle spatial simple. Pour surmonter ces incompatibilités et générer une suite de grilles d'entrée entièrement standardisées dans un format géographique défini de manière identique, une chaîne de traitement a été développée avec les étapes suivantes :

1. Chaque source de données d'entrée a été re-projetée, si nécessaire, en utilisant une projection de plaque carrée équirectangulaire standardisée sous le système de coordonnées 1984 du Système Géodésique Mondial.
2. Les grilles d'entrée définies à différentes résolutions spatiales ont été ré-échantillonnées à  $5 \times 5$  km.
3. Les grilles ont été étendues ou découpées pour correspondre à une étendue standardisée du pays.
4. Un algorithme sur mesure a été élaboré pour comparer chaque grille rectifiée et ré-échantillonnée à un format de référence démarquant les contours terre-mer pour chaque pays. Ceci a utilisé une procédure d'interpolation simple et/ou de découpage pour aligner les nouvelles grilles sur ce modèle de référence, ce qui a permis de s'assurer que toutes les lignes littorales étaient parfaitement cohérentes pixel par pixel.

Les covariables géospatiales peuvent être statiques (à un moment donné dans le temps), multi-temporelles (plusieurs couches spatiales représentant plusieurs points régulièrement espacés dans le temps) ou synoptiques (sur une longue période et résumées pour montrer une moyenne de long terme ou d'autres tendances générales). La période de temps (date) que représentent les covariables varie et ne correspond pas nécessairement à la période exacte de l'enquête DHS qui est en train d'être modélisée. Les limites possibles des covariables dans le modèle sont discutées plus en détail dans la section 4.1.2.

**Tableau 2. Résumé des covariables utilisées dans la modélisation des données publiées en septembre 2016**

Nom court	Description	Source de données originale	Temporalité	Date
<b>Access</b>	Temps de voyage vers villes > 50k via tous moyens de transport	<a href="http://forobs.jrc.ec.europa.eu">http://forobs.jrc.ec.europa.eu</a>	Statique	2000
<b>Aridity</b>	Aridité annuelle moyenne	<a href="http://csi.cgiar.org/Aridity/">http://csi.cgiar.org/Aridity/</a>	Synoptique	1950–2000
<b>NTL</b>	Lumières de nuit VIIRS –2012	<a href="http://ngdc.noaa.gov/eog/">http://ngdc.noaa.gov/eog/</a>	Statique	2012
<b>Elevation</b>	Modèles d'élévation numérique quasi-globaux (DEMs) SRTM	<a href="http://webmap.ornl.gov/">http://webmap.ornl.gov/</a>	Statique	2000
<b>EVI</b>	Indice de végétation amélioré	<a href="http://modis.gsfc.nasa.gov/">http://modis.gsfc.nasa.gov/</a>	Multitemporel	2001–2014
<b>GPW</b>	Densité de population GPW	<a href="http://sedac.ciesin.columbia.edu/">http://sedac.ciesin.columbia.edu/</a>	Statique	2010
<b>LST.day</b>	Température de la surface du sol pendant la journée	<a href="http://modis.gsfc.nasa.gov/">http://modis.gsfc.nasa.gov/</a>	Multitemporel	2001–2014
<b>LST.delta</b>	Plage de fluctuation quotidienne de la température de surface du sol	<a href="http://modis.gsfc.nasa.gov/">http://modis.gsfc.nasa.gov/</a>	Multitemporel	2001–2014
<b>LST.night</b>	Température de la surface du sol pendant la nuit	<a href="http://modis.gsfc.nasa.gov/">http://modis.gsfc.nasa.gov/</a>	Multitemporel	2001–2014
<b>PET</b>	Évapotranspiration potentielle annuelle moyenne	<a href="http://csi.cgiar.org/Aridity/">http://csi.cgiar.org/Aridity/</a>	Synoptique	1950–2000
<b>PRECIP</b>	Moyenne des précipitations mensuelles	<a href="http://www.worldclim.org/">http://www.worldclim.org/</a>	Synoptique	1950–2000
<b>TCB</b>	Luminosité de Tasseled-cap	<a href="http://modis.gsfc.nasa.gov/">http://modis.gsfc.nasa.gov/</a>	Multitemporel	2001–2014
<b>TCW</b>	Humidité de Tasseled-cap	<a href="http://modis.gsfc.nasa.gov/">http://modis.gsfc.nasa.gov/</a>	Multitemporel	2001–2014

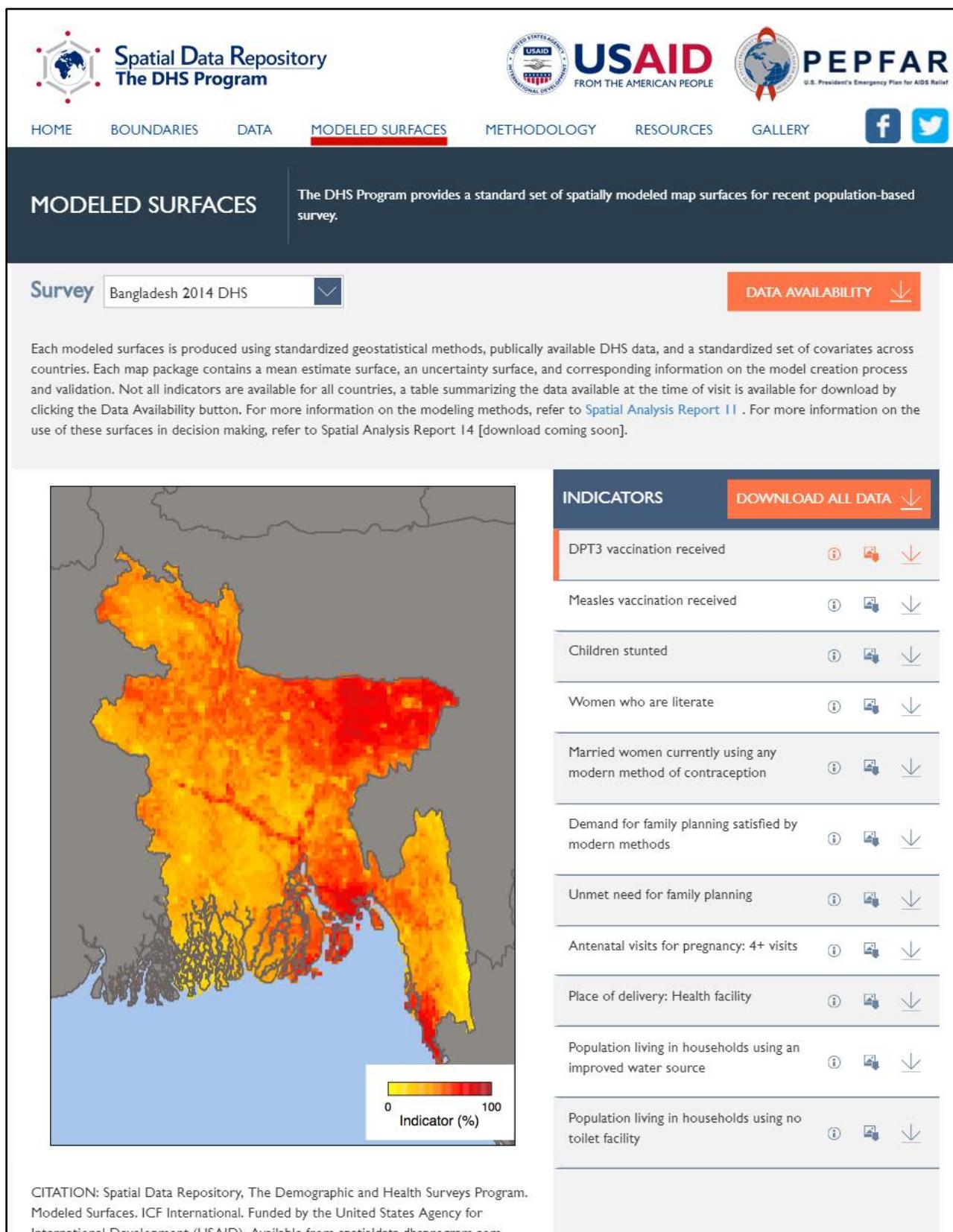
## 2.2 Disponibilité des surfaces modélisées

Les surfaces modélisées sont publiquement disponibles et peuvent être téléchargées à partir du Répertoire de Données Spatiales du DHS Program (<http://spatialdata.dhsprogram.com/>). La Figure 3 montre le site Web. Un tableau résumant les données disponibles au moment de la visite peut être téléchargé sur le site; pour chaque pays, il indique l'année de l'enquête disponible et les indicateurs sont disponibles pour chacune de ces enquêtes.

Le package de téléchargement est un dossier compressé contenant 5 composants :

- Surface modélisée de l'estimation ponctuelle à une résolution de  $5 \times 5$  km (format GeoTIFF)
- Surface modélisée de l'estimation de l'incertitude à une résolution de  $5 \times 5$  km (format GeoTIFF)
- Image de la surface modélisée de l'estimation moyenne (format PNG)
- Image de la surface modélisée de l'estimation de l'incertitude (format PNG)
- Document spécifique sur les procédures de modélisation (format PDF)

Figure 3. Capture d'écran du site de téléchargement de surfaces modélisées



## 2.2.1 Ensemble des surfaces modelisées et conventions de dénomination des fichiers associés

Les conventions de dénomination du package des surfaces modélisées suivent une norme dérivée de l'application API (Interface de Programmation Applicative, ou « Application Program Interface ») du DHS Program. L'application API fournit des noms standard aux enquêtes et aux indicateurs par pays, et permet aux utilisateurs de trouver les informations relatives à l'enquête et à l'indicateur. Chaque fichier de données a une convention de dénomination standard qui identifie le pays, l'année d'enquête, l'indicateur, le type et le numéro de la version du fichier de données. Les champs sont décrits ci-dessous.

- Champ 1: SurveyID (à partir de l'API du DHS Program, voir <http://api.dhsprogram.com/rest/dhs/surveys?f=html>)
- Champ 2: SDRID (Forme courte de l'Id de l'indicateur à partir de l'API du DHS Program, voir <http://api.dhsprogram.com/ rest/dhs/indicators?returnFields=IndicatorId,SDRID,Label,Definition&f=html>)
- Champ 3: MS (surfaces modélisées)
- Champ 4: TYPE (soit MEAN pour estimation ponctuelle, soit CI pour estimation de l'incertitude)
- Champ 5: v# (Le numéro de version du fichier de données devrait être "v01" dans la plupart des cas, sauf si le fichier de données a été réédité).

Ces champs sont combinés pour chaque composante du package des données. Ils sont décrits dans le Tableau 3 avec un exemple pour l'enquête DHS 2008 du Ghana et l'indicateur de retard de croissance chez les enfants de moins de 5 ans.

**Tableau 3. Convention de nomination des surfaces modelisées**

	Générique	Exemple
<b>Nom de dossier</b>	SurveyID _SDRID_MS_v#	GH2008DHS_CNNUTSCHA2_MS_v01
<b>Fichiers de données</b>	SurveyID _SDRID_MS_TYPE_v#	GH2008DHS_CNNUTSCHA2_MS_MEAN_v01 GH2008DHS_CNNUTSCHA2_MS_CI_v01
<b>Fichiers image</b>	SurveyID _SDRID_MS_TYPE_v#	GH2008DHS_CNNUTSCHA2_MS_MEAN_v01 GH2008DHS_CNNUTSCHA2_MS_CI_v01
<b>Documentation</b>	SurveyID _SDRID_MS_v#	GH2008DHS_CNNUTSCHA2_MS_v01

## 2.2.2 Attribution

Les fichiers de données de surface modélisées du DHS Program sont publiquement et gratuitement disponibles. Vous devez présenter une citation appropriée lorsque vous utilisez les fichiers de données de surfaces modélisées du DHS Program. Les utilisateurs de données doivent citer le Répertoire des Données Spatiales comme source de toutes les analyses, rapports, publications, présentations et autres produits dérivés. Pour utiliser la citation recommandée, remplacez simplement la date d'accès par la date réelle de téléchargement.

Pour un seul fichier de données de surface modélisé:

*Répertoire de Données Spatiales, The Demographic and Health Surveys Program. Surfaces modelisées. SurveyID\_SDR-API ID\_MS v #. ICF. Financé par l'Agence américaine pour le développement international (USAID). Disponible sur spatialdata.dhsprogram.com. [Visité le JOUR MOIS ANNEE]*

Pour plusieurs fichiers de données de surface modélisée:

*Répertoire de Données Spatiales, The Demographic and Health Surveys Program. Surfaces modelisées. ICF. Financé par l'Agence américaine pour le développement international (USAID). Disponible sur spatialdata.dhsprogram.com. [Visité le JOUR MOIS ANNEE]*



### 3 Comment les surfaces modélisées du DHS Program sont-elles créées?

#### Résumé

Cette section décrit la préparation des données des grappes géoréférencées pour chaque indicateur DHS sélectionné, l'assemblage et l'exploration d'une série de covariables géospatiales quadrillées et l'utilisation de ces données dans une série de modèles géostatistiques bayésiens adaptés pour générer des surfaces modélisées finales correspondant à chaque indicateur. Les résultats des modèles sont des surfaces modélisées à une résolution de pixel de  $5 \times 5$  km. Une description plus détaillée du processus de modélisation est présentée dans le RAS 11 (Gething et al., 2015).

#### Questions clés

##### *Puis-je reproduire moi-même ces surfaces modélisées ?*

Tous les détails sur les entrées du modèle (covariables) et la structure générale du modèle sont partagés dans ce document et dans d'autres documents d'accompagnement. En théorie, muni des bonnes compétences, vous pouvez créer vos propres surfaces modélisées; cependant, vos surfaces peuvent être différentes des surfaces du DHS Program car certaines décisions de modélisation ont lieu dans le cycle du modèle et peuvent modifier le résultat final. Voir la section 3.1.

##### *Quelles covariables forment chaque carte ?*

Toutes les covariables sont utilisées en entrée pour chaque carte. Cependant, une procédure d'ajustement est utilisée pour pondérer automatiquement l'influence de chaque covariable en fonction de la quantité d'informations utiles qu'elle contient sur l'indicateur d'intérêt. La documentation associée à chaque surface décrit la contribution relative de chaque covariable dans le modèle ajusté final (exprimée en pourcentage). Voir la section 3.1.3.

##### *Pourquoi les zones ont-elles des niveaux d'erreur différents ?*

L'erreur dans ce contexte se réfère à la capacité du modèle à prédire la valeur correcte d'un indicateur dans une localité particulière. Cela dépend de plusieurs facteurs comme le nombre de grappes dans l'enquête, la densité des grappes de l'enquête autour d'un emplacement de prédiction, le nombre de répondants dans chaque grappe, la force des corrélations entre les covariables et l'indicateur, et le degré inhérent de la variation spatiale affichée par l'indicateur. La capacité prédictive globale du modèle est résumée dans la statistique de validation de l'erreur absolue moyenne. Voir la section 3.3.

##### *Les valeurs d'estimation moyenne des indicateurs et les cartes d'incertitude peuvent-elles être affichées simultanément ?*

Oui, il existe plusieurs façons d'afficher les données simultanément. Voir la section 3.3.

### 3.1 Explication du processus de création de surfaces modélisées

La figure 4 illustre le processus de création de surfaces modélisées du DHS Program, à partir des entrées du modèle jusqu'aux sorties du modèle. Le processus est, en outre, décrit dans les sections suivantes pour chaque étape numérotée dans la figure.

Les détails de l'approche sont expliqués dans le Rapport d'Analyse Spatiale EDS 11 (Gething et al., 2015). Une approche de modélisation géostatistique bayésienne (MBG) (Diggle et Ribeiro 2007, Diggle, Tawn et Moyeed 1998) a été utilisée pour générer les surfaces modélisées. S'appuyant sur des techniques initialement conçues pour une cartographie détaillée de la prévalence du paludisme (Gething et al 2011, Hay et al., 2009), les modèles MBG représentent la variation observée dans les données d'enquête au niveau des grappes en utilisant quatre composantes.

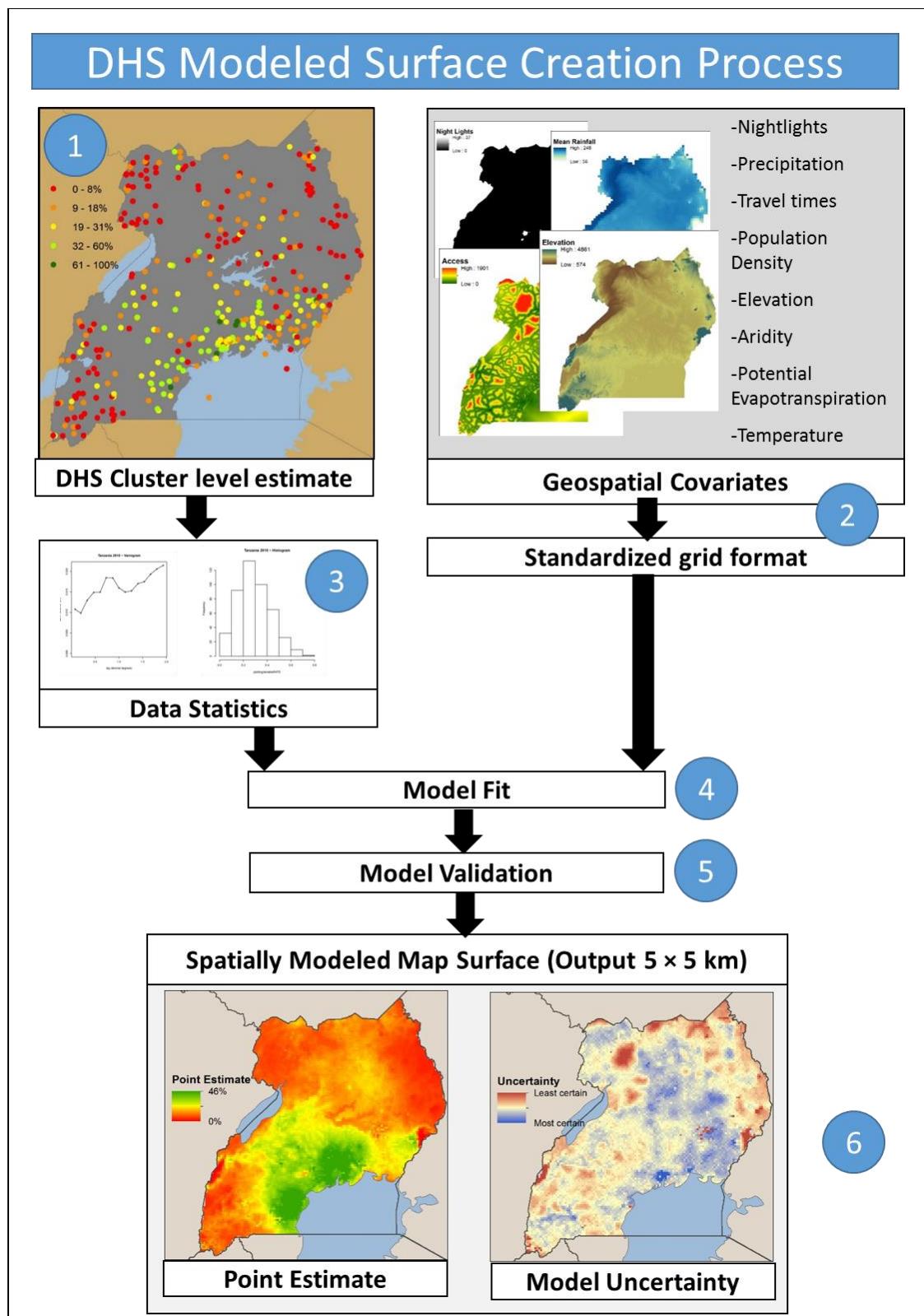
- L'erreur d'échantillonnage, qui peut souvent être importante étant donné la petite taille des échantillons dans les grappes individuelles, est représentée par un modèle d'échantillonnage standard; généralement de type binomial lorsque l'indicateur en question est une proportion.
- Certaines variations non dues à l'échantillonnage peuvent être expliquées en utilisant des effets fixes, par une relation de régression multivariée définie en liant la variable-indicateur à une série de covariables géospatiales.
- Les erreurs supplémentaires non dues à l'échantillonnage qui ne sont pas expliquées par les effets fixes sont généralement auto-corrélées spatialement, et sont représentées en utilisant une composante d'effets aléatoires. Une distribution spatiale normale multivariée connue sous le nom de processus gaussien est utilisée et paramétrée par une fonction de covariance spatiale.
- Toute variation restante non capturée par ces composantes est représentée par un simple terme de bruit gaussien équivalent à celui employé dans un modèle linéaire non spatial standard.

#### 3.1.1 Entrées du modèle (étapes 1 et 2)

Deux types de données sont utilisés dans le processus de modélisation de surface (étapes 1 et 2 de la figure 4).

1. Les observations à l'échelle de la grappe de la DHS : le numérateur et le dénominateur de l'indicateur au niveau de la grappe sont créés avec les données EDS publiquement disponibles (fichiers recodés des individus et des ménages). Cette information est ensuite liée aux données de localisation GPS au niveau de la grappe.
2. Les covariables géospatiales : une gamme de covariables est incluse en tant que covariables explicatives possibles. Un aspect important de la modélisation géostatistique est l'exploitation des covariables géospatiales qui sont en corrélation avec la variable dépendante, qui peuvent expliquer, en partie, la variation de cette variable et permettre des prédictions plus précises à travers la carte. Comme décrit ci-dessus, une suite de covariables a été choisie dans le répertoire existant de l'Université d'Oxford sur la base de facteurs dont la correlation avec les indicateurs démographiques et de santé dans différents contexts a été préalablement démontrée. Les covariables sont standardisées suivant une grille de rasters de  $5 \times 5$  km sur un littoral uniforme.

Figure 4. Diagramme schématique résumant le processus de création de surfaces modélisées des données DHS

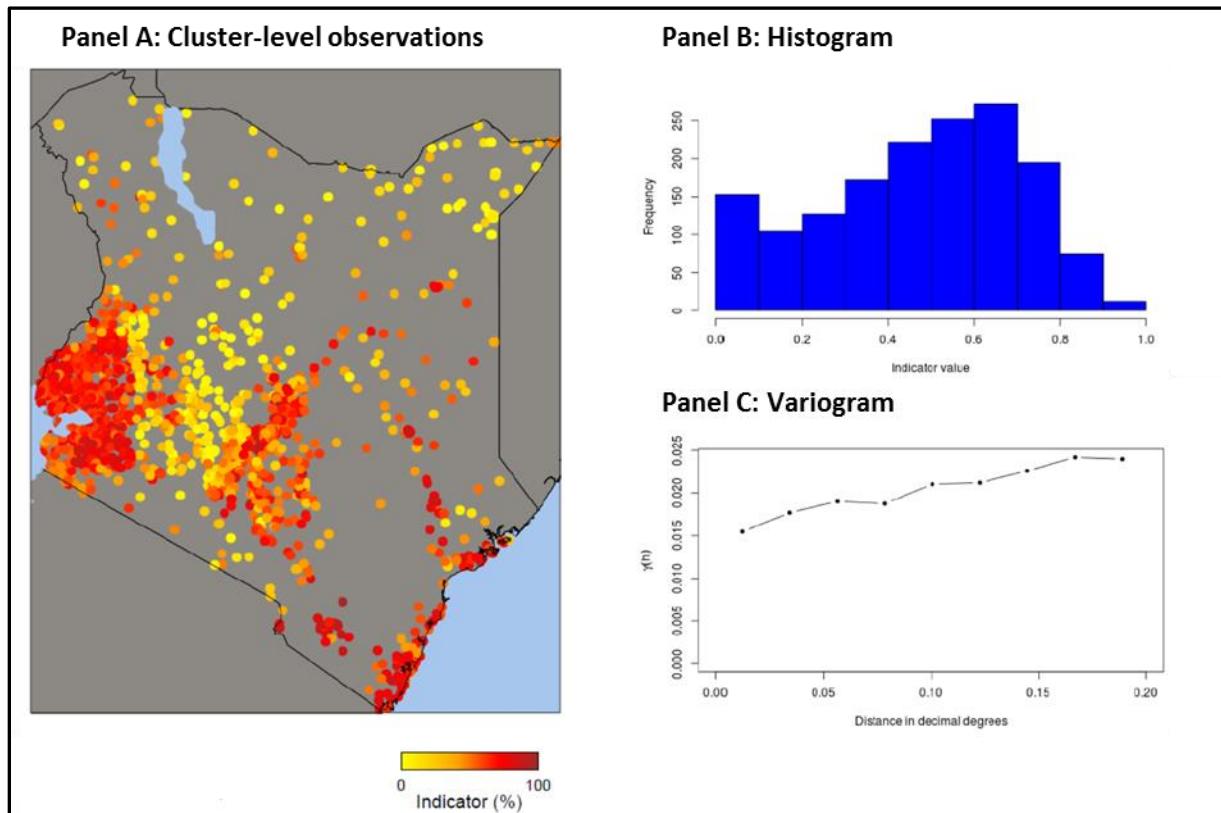


### 3.1.2 Statistiques des données (étape 3)

Trois analyses de base ont exploré les caractéristiques des données brutes (étape 3 de la figure 4).

- **Observations à l'échelle de la grappe:** une carte montrant l'emplacement et les valeurs observées de l'indicateur pour chaque grappe géo-localisée de l'enquête DHS, chaque grappe étant représentée par un point (Figure 5, encadré A).
- **Histogramme:** un histogramme empirique simple pour évaluer la distribution statistique, ce qui est utile pour interpréter l'indicateur et les cartes résultantes (Figure 5, encadré B). Les histogrammes avec des valeurs représentées sur un grand nombre de différentes valeurs de l'indicateur produiront une carte avec un spectre de valeurs alors que les histogrammes où la majorité des valeurs sont regroupées produiront une carte avec un aspect plus uniforme.
- **Variogramme:** La structure d'autocorrélation spatiale présente dans les grappes est évaluée par un variogramme empirique (Figure 5, encadré C). Un variogramme trace la semi-variance (la dissimilarité moyenne des valeurs de l'indicateur entre deux points de grappe) par rapport au décalage spatial (la distance géographique séparant 2 points). Lorsque les données sont structurées spatialement, la forme caractéristique du variogramme montrera une semi-variance en augmentation constante avec le décalage. Inversement, les données sans structure spatiale conduisent à un variogramme plat. Les variables avec une plus grande autocorrélation spatiale tendent à se prêter davantage à l'interpolation spatiale et à des cartes plus fiables.

Figure 5. Exemple de statistiques de données



### 3.1.3 Ajustement du modèle (étape 4)

Après avoir considéré les statistiques préliminaires des données, l'étape suivante du processus de modélisation consiste à paramétriser le modèle pour déterminer l'ensemble des valeurs des paramètres du modèle qui permettent le meilleur ajustement possible aux données (étape 4 de la Figure 4). Les modèles MBG comportent trois catégories de paramètres, ce qui nécessite d'établir les caractéristiques du processus gaussien (effets aléatoires), la nature et l'ampleur de la contribution de chaque covariable (effets fixes) et l'erreur résiduelle non corrélée (la composante non spatiale). Le processus gaussien spatial est gouverné par une fonction de covariance spatiale (la fonction Matern a été utilisée pour sa flexibilité) qui est paramétrée par un paramètre d'échelle (qui caractérise la distance spatiale sur laquelle les points sont autocorrélés) et un paramètre de variance (qui décide l'ampleur de l'autocorrelation) (voir le Tableau 4, Paramètres du modèle). Les effets fixes ont une structure de paramètres plus complexe qui permet des relations non-linéaires entre les covariables et la variable réponse ainsi que des interactions entre elles. Une approche de "régularisation" permet d'utiliser la suite complète des covariables dans le modèle sans risque de surapprentissage. Toutes les covariables restent dans le modèle bien que leurs contributions relatives aux prédictions finales puissent être considérables ou presque nulles, afin de maximiser la performance prédictive. Ces contributions sont décrites comme une contribution par covariable en pourcentage (voir le Tableau 4, Contributions des covariables). Cette approche diffère légèrement de l'approche décrite dans le RAS 11 dans lequel il y avait un processus initial de sélection de covariables avec seulement les covariables sélectionnées incluses dans le modèle final. La nouvelle approche fournit une performance constamment plus élevée et présente l'avantage supplémentaire de réduire l'apport subjectif d'un analyste en affinant la sélection des covariables. L'erreur résiduelle non corrélée est paramétrée par un seul paramètre de variance (voir le Tableau 4, Paramètres du modèle). Tous les paramètres sont estimés conjointement dans un seul processus d'ajustement en utilisant l'inférence bayésienne avec des paramètres à priori vagues.

**Tableau 4. Exemple de paramètres du modèle, de contribution des covariables et de statistiques de validation**

Paramètres du modèle	
<b>Fonction de covariance</b>	Matern
<b>Echelle de corrélation spatiale (km)</b>	70.9
<b>Variance de la composante spatiallement structurée</b>	0.5
<b>Variance de la composante non spatiale</b>	0.58
Contributions des covariables	
<b>Aridité annuelle moyenne</b>	12%
<b>Lumière de nuit VIIRS –2012</b>	8%
<b>Indice de végétation amélioré</b>	4%
<b>Modèles d'élévation numérique quasi-globaux (DEMs) SRTM</b>	14%
<b>Densité de la population GPW</b>	4%
<b>Température de la surface du sol pendant la journée</b>	18%
<b>Plage de fluctuation quotidienne de la température de surface du sol</b>	4%
<b>Température de la surface du sol pendant la nuit</b>	6%
<b>Évapotranspiration potentielle annuelle moyenne</b>	6%
<b>Luminosité de Tasseled-cap</b>	6%
<b>Humidité de Tasseled-cap</b>	9%
<b>Temps de voyage vers villes &gt; 50k via toutes méthodes de transport</b>	6%
<b>Moyenne des précipitations mensuelles</b>	6%

Statistiques de validation	
Corrélation	88%
Erreur Absolue Moyenne	9%
Erreur Quadratique Moyenne	1%

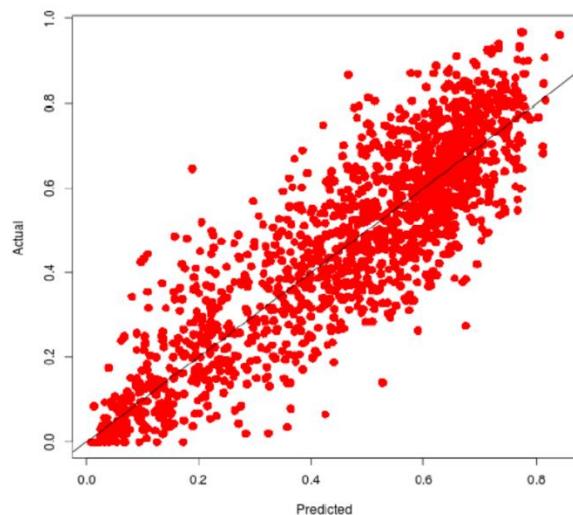
### 3.1.4 Validation du modèle (étape 5)

La validation du modèle et les statistiques correspondantes constituent une mesure importante de la performance prédictive du modèle géostatistique (étape 5 dans la Figure 4). La performance est évaluée par une validation hors-échantillon qui comprend une procédure de suspension à quatre volets dans laquelle 25 % des points de données sont retirés de façon aléatoire de l'ensemble de données. Le modèle est ensuite exécuté intégralement avec les 75 % de points de données restants et les valeurs prédites aux emplacements des données retirées sont comparées à leurs valeurs observées. Ce processus est répété quatre fois sans remplacement, de sorte que chaque point de données est retiré une fois pendant les quatre cycles de validation. Les statistiques de validation standard sont ensuite calculées en tant que mesures de la précision du modèle:

- Degré d'association linéaire entre les valeurs observées et prédites (corrélation, COR).
- Erreur Absolue Moyenne (EAM) qui quantifie la précision du modèle, qui est l'amplitude moyenne de la différence entre les valeurs observées et prédites. Ceci est calculé dans les mêmes unités que la variable prédite; par exemple, si l'indicateur est un taux exprimé sur une échelle de 0 à 100 %, l'EAM sera également une valeur comprise entre 0 et 100 %.
- L'Erreur Quadratique Moyenne (EQM) qui indique l'adéquation du modèle et encapsule le biais et l'erreur, avec des valeurs proches de zero indiquant que le modèle est plus précis et proche de un, moins précis.

Des exemples de ces statistiques sont présentés dans le Tableau 4 (Statistiques de validation). La Figure 6 ci-dessous montre un diagramme de dispersion avec la distribution des points réels (observés) par rapport aux points prédits du modèle.

**Figure 6. Exemple de diagramme de dispersion de validation**

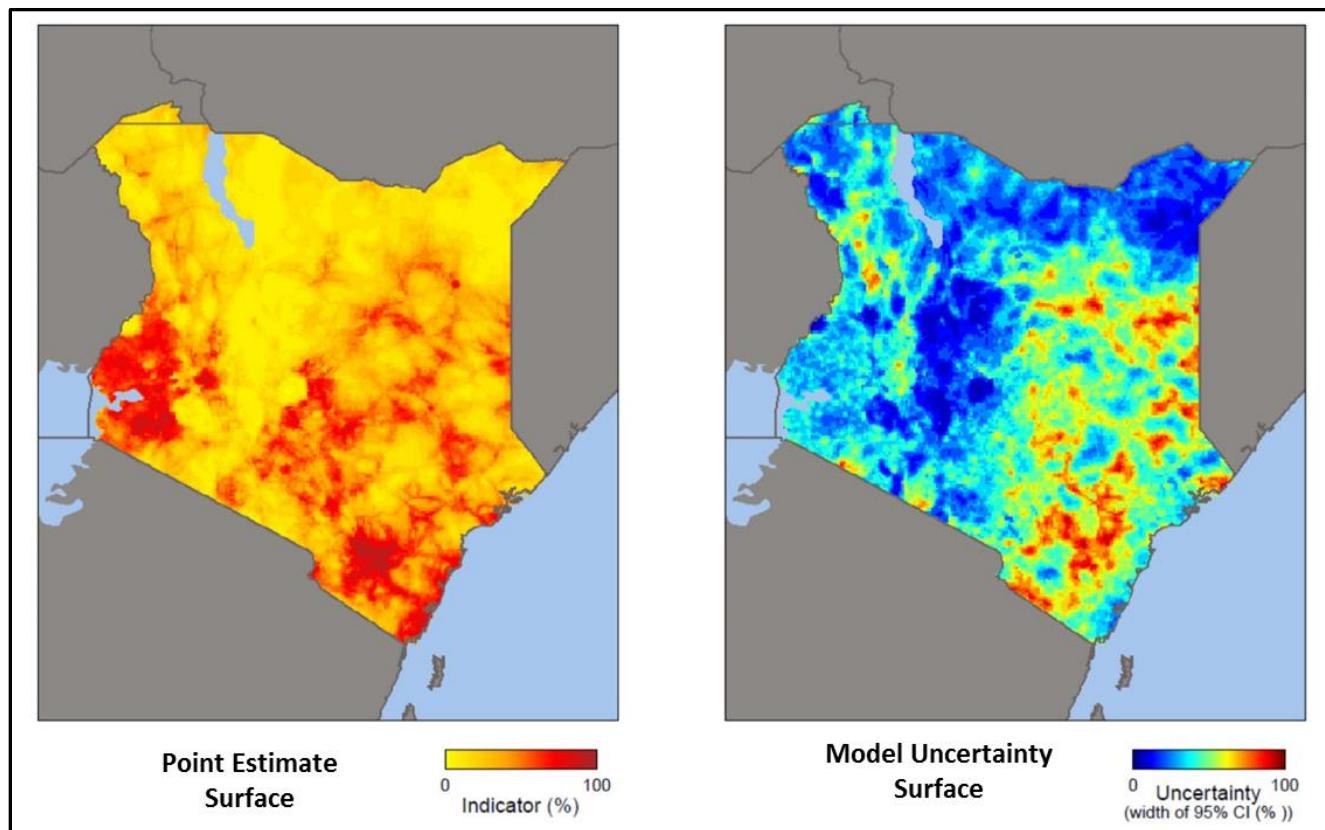


### 3.1.5 Sorties du modèle (étape 6)

Dans le package de téléchargement, le modèle a deux cartes de surface comme sortie avec une résolution de pixel de  $5 \times 5$  km (étape 6 sur la Figure 4): la surface de l'estimation ponctuelle et la surface de l'incertitude du modèle. Les modèles MBG génèrent des estimations de la variable d'intérêt à chaque emplacement sur une surface quadrillée. Pour chacun de ces pixels, la sortie du modèle complet est une distribution à posteriori pour l'indicateur prédict qui présente un modèle complet de l'incertitude autour de la valeur estimée. Ceux-ci peuvent être résumés avec une estimation ponctuelle (telle que la moyenne postérieure) pour générer une surface cartographiée. Des statistiques sommaires supplémentaires de chaque distribution postérieure peuvent ensuite être cartographiées pour illustrer le degré d'incertitude associé à chaque valeur prédict.

- **Surface de l'estimation ponctuelle:** La carte trace la valeur d'estimation ponctuelle modélisée pour chaque pixel de  $5 \times 5$  km sur la base de données géo-localisées à l'échelle de la grappe issues de l'enquête. Cette valeur représente effectivement la valeur attendue de l'indicateur dans cette région de  $5 \times 5$  km. Les indicateurs modélisés étant des taux (prévalence ou proportion), tous les taux se situent sur une échelle comprise entre 0 et 1 (ou 0 % et 100 %) (Figure 7, à gauche).
- **Surface de l'incertitude du modèle:** Une carte d'incertitude associée résume le niveau de certitude associé aux valeurs indiquées dans la carte de l'estimation ponctuelle en affichant la largeur complète de l'intervalle crédible à 95 % (également appelé intervalle de confiance ou IC) pour chaque pixel (Figure 7, à droite). Dans une situation d'incertitude totale sur la valeur d'un pixel, l'IC à 95 % couvrirait toute la plage et la valeur réelle pourrait se situer n'importe où entre zéro et un. Inversement, lorsqu'une variable est prédict avec une très grande certitude, la largeur de l'IC à 95 % peut être très étroite. En d'autres termes, il y a une probabilité de 95 % que la vraie valeur se situe dans une fourchette étroite de valeurs possibles; cela indique que la prédition a une faible incertitude.

**Figure 7. Exemple de surface de l'estimation ponctuelle et de surface de l'incertitude du modèle**



### 3.2 Exemple de documentation de surface modélisée

Les pages suivantes décrivent un exemple de documentation de surface modélisée incluse dans chaque fichier compressé qui peut être téléchargé du site web du Répertoire des Données Spatiales du DHS Program. Chaque indicateur de chaque enquête de chaque pays a son propre document avec les figures et les informations pertinentes pour cette surface modelisée. Le texte principal, qui est standard dans tous les documents, aide les utilisateurs à comprendre les informations dans les figures.

## Persons with access to an ITN

### Kenya 2014 Demographic and Health Survey

#### **The DHS Program**

The Demographic and Health Survey (DHS) Program has been a leader in collecting and providing cluster-randomized survey data on core development indicators since 1984. DHS surveys are configured to provide indicator estimates at the national and in more recent surveys first level administrative level. In more recent years, the availability of the Global Positioning System (GPS) coordinates for DHS household survey clusters provides highly local scale information that can be linked with survey outputs for quantifying demographic and health status heterogeneities and inequities. In 2016, the DHS Program started publically providing a standard set of spatially modeled map surfaces. The surfaces use publicly available spatial covariate data, are standardized across countries, and are comparable in order to facilitate policy and program decision-making at levels below the current survey representative sub-national areas. The DHS Program is funded by the U.S. Agency for International Development (USAID) (<http://dhsprogram.com/>).

#### **Indicator definition: Persons with access to an ITN**

**Full definition:** Percentage of the de facto household population who could sleep under an ITN if each ITN in the household were used by up to two people

**Numerator:** De facto household population who could sleep under an ITN if each ITN in the household were used by up to two people

**Denominator:** De facto household population

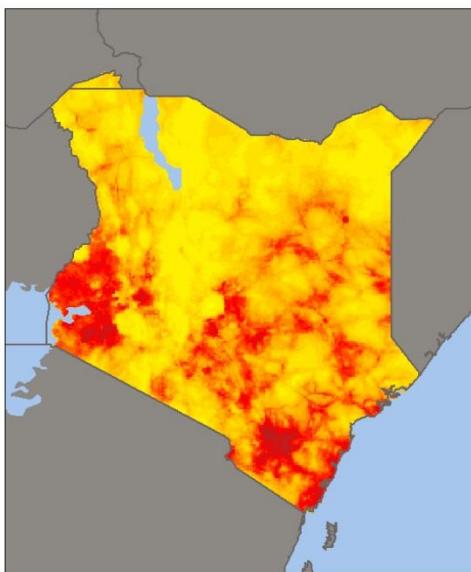
**API ID:** ML\_ITNA\_P\_ACC

#### **Geospatial modelling**

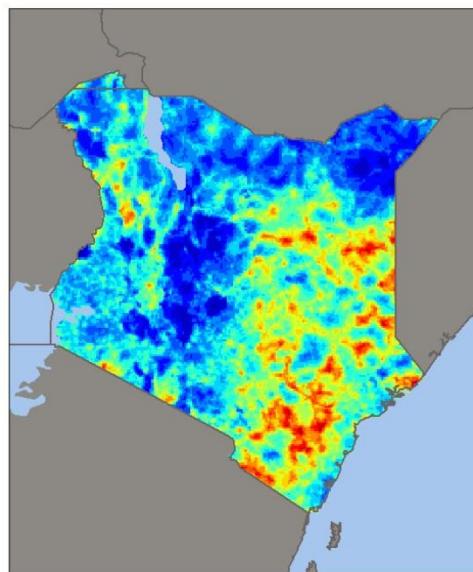
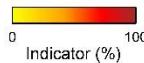
A Bayesian model-based geostatistical (MBG) approach (Diggle and Ribeiro 2007, Diggle, Tawn, and Moyeed 1998) was used to generate the interpolated surfaces shown below. Building on techniques originally conceived for detailed mapping of malaria prevalence (Hay et al., 2009, Gething et al 2011), MBG models represent the observed variation in cluster-level survey data using four components. (i) Sampling error, which can often be large given the small sample sizes in individual clusters, is represented using a standard sampling model, usually the binomial. (ii) Some non-sampling variation can often be explained using fixed effects, whereby a multivariate regression relationship is defined linking the indicator variable with a suite of geospatial covariates. (iii) Additional non-sampling errors not explained by the fixed effects are usually spatially autocorrelated, and this is represented using a random effects component. A spatial multi-variate normal distribution known as a Gaussian Process is employed, parameterized by a spatial covariance function. (iv). Finally, any remaining variation not captured by these components is represented using a simple Gaussian noise term equivalent to that employed in a standard non-spatial linear model. This approach is explained in full in the DHS Spatial Analysis Reports No. 11 (Gething et al, 2015).

#### **Maps**

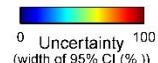
MBG models generate estimates of the variable of interest at each location on a gridded surface. The maps below were defined on a grid where each individual pixel measures approximately 5x5 km. The full model output is, for each of those pixels, a posterior distribution for the predicted indicator, representing a complete model of the uncertainty around the estimated value. These can be summarized using a point estimate (such as the posterior mean) to generate a mapped surface (as shown in Figure 1). Additional summary statistics from each posterior distribution can also be mapped to illustrate the degree of uncertainty associated with each predicted value. Figure 2 shows one such metric of uncertainty – the width of the 95% credible intervals, with large values representing areas of the highest uncertainty, and vice versa.



**Figure 1. Interpolated surface for the indicator.**  
 The map plots the point estimate for each 5x5 km pixel based on geo-located cluster-level data from the survey.



**Figure 2. Uncertainty surface for the indicator.**  
 The map plots the uncertainty for each pixel, measured using the width of the 95% credible intervals.



This document can be cited as follows: Spatial Data Repository, The Demographic and Health Surveys Program. Modeled Surfaces: KE2014DHS\_MLITNAPACC\_MS\_v01  
 ICF International. Funded by the United States Agency for International Development (USAID). Available from [spatialdata.dhsprogram.com](http://spatialdata.dhsprogram.com) [Accessed DAY MONTH YEAR]

The information provided on this sheet is not official U.S. Government information and does not represent the views or positions of the U.S. Agency for International Development or the U.S. Government

## Persons with access to an ITN Kenya 2014 Demographic and Health Survey

### Data statistics

Figure 3 shows the location and observed values of the indicator for each geo-located DHS survey cluster. Two basic exploratory analyses were undertaken to explore the characteristics of these raw data. First, a simple empirical histogram were generated to assess statistical distribution which can be useful in interpreting the indicator and resulting maps. This is shown in Figure 4 (top). Second, the spatial autocorrelation structure present in the indicator values across clusters was assessed via an empirical variogram (Figure 4 (bottom)). A variogram plots semivariance (the average dissimilarity in the indicator values between two cluster points) against spatial lag (the geographical distance separating two points). Where data are spatially structured, a characteristic variogram form is for semivariance to steadily increase with increasing lag. Conversely, data with no spatial structure lead to a flat variogram. Variables with greater spatial autocorrelation tend to be more amenable to spatial interpolation and more reliable maps.

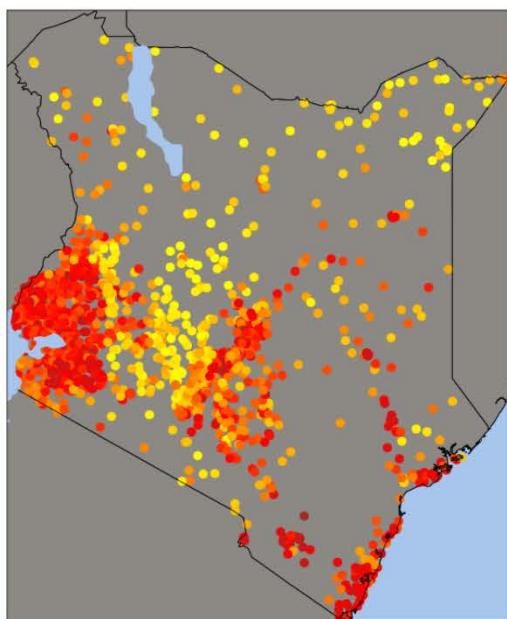


Figure 3. Cluster-level observations of the indicator. These geo-located data formed the basis for the modeled interpolated surface.

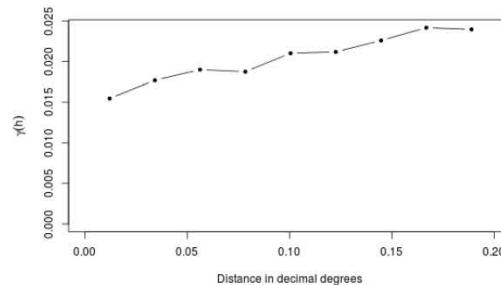
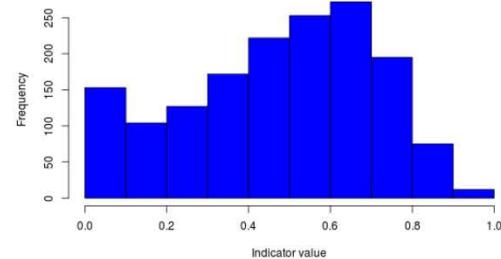


Figure 4. Histogram (top) and variogram (bottom) of the cluster-level indicator data from the survey.

### Geospatial covariates

An important aspect of geostatistical modeling is the exploitation of geospatial covariates that are correlated with the indicator of interest, and thus partially explain variation in that indicator, allowing more accurate predictions across the map. A suite of geospatial covariates were chosen from existing libraries that have previously demonstrated broad utility in geospatial mapping (Weiss et al 2014, Gething et al, 2015) – listed in full in Table 1. A ‘regularization’ approach was used which allows a large suite of covariates to be used in the model without risking over-fitting: all covariates remain in the model but their relative contributions to the final predictions can be large or almost zero in order to maximize predictive performance. Table 2 describes the relative contributions of each covariate to the predictions presented here.

Short name	Description	Original Data Source	Temporal	Date
Access	Travel time to cities with > 50k via all transport methods	<a href="http://forobs.jrc.ec.europa.eu">http://forobs.jrc.ec.europa.eu</a>	Static	2000
Aridity	Mean annual aridity	<a href="http://csi.cgiar.org/Aridity/">http://csi.cgiar.org/Aridity/</a>	Synoptic	1950–2000
NTL	VIIRS Nighttime Lights–2012	<a href="http://ngdc.noaa.gov/eog/">http://ngdc.noaa.gov/eog/</a>	Static	2012
Elevation	SRTM Near-global Digital Elevation Models (DEMs)	<a href="http://webmap.ornl.gov/">http://webmap.ornl.gov/</a>	Static	2000
EVI	Enhanced vegetation index	<a href="http://modis.gsfc.nasa.gov/">http://modis.gsfc.nasa.gov/</a>	Multitemporal	2001–2014
GPW	GPW population density	<a href="http://sedac.ciesin.columbia.edu/">http://sedac.ciesin.columbia.edu/</a>	Static	2010
LST.day	Land surface temperature in the daytime	<a href="http://modis.gsfc.nasa.gov/">http://modis.gsfc.nasa.gov/</a>	Multitemporal	2001–2014
LST.delta	Land surface temperature daily fluctuation range	<a href="http://modis.gsfc.nasa.gov/">http://modis.gsfc.nasa.gov/</a>	Multitemporal	2001–2014
LST.night	Land surface temperature in the nighttime	<a href="http://modis.gsfc.nasa.gov/">http://modis.gsfc.nasa.gov/</a>	Multitemporal	2001–2014
PET	Mean annual Potential Evapotranspiration	<a href="http://csi.cgiar.org/Aridity/">http://csi.cgiar.org/Aridity/</a>	Synoptic	1950–2000
PRECIP	Average monthly rainfall	<a href="http://www.worldclim.org/">http://www.worldclim.org/</a>	Synoptic	1950–2000
TCB	Tasseled-cap brightness	<a href="http://modis.gsfc.nasa.gov/">http://modis.gsfc.nasa.gov/</a>	Multitemporal	2001–2014
TCW	Tasseled-cap wetness	<a href="http://modis.gsfc.nasa.gov/">http://modis.gsfc.nasa.gov/</a>	Multitemporal	2001–2014

Table 1. Further details of geospatial covariates included in the model

This document can be cited as follows: Spatial Data Repository, The Demographic and Health Surveys Program. Modeled Surfaces. KE2014DHS\_MLITNAPACC\_MS\_v01  
ICF International. Funded by the United States Agency for International Development (USAID). Available from [spatialdata.dhsprogram.com](http://spatialdata.dhsprogram.com). [Accessed DAY MONTH YEAR]

The information provided on this sheet is not official U.S. Government information and does not represent the views or positions of the U.S. Agency for International Development or the U.S. Government



## Persons with access to an ITN Kenya 2014 Demographic and Health Survey

### Model fit and validation

The predictive performance of the geostatistical model was assessed via out-of-sample validation. This consisted of a four-fold hold-out procedure whereby 25% of the data points were randomly withdrawn from the dataset, the model run in full using the remaining 75% of data, and the predicted values at the locations of the hold-out data compared to their observed values. This was repeated four times without replacement such that every data point was held out once across the four validation runs. Standard validation statistics were computed as measures of model precision (mean absolute error, MAE), bias (mean square error, MSE), and the degree of linear association between observed and predicted values (correlation, COR). The MAE quantifies model precision, while the MSE indicates how biased the model is, with values close to zero providing an indication that the model is unbiased. These statistics are listed in Table 2, and a scatterplot showing the distribution of observed versus predicted points in Figure 5.

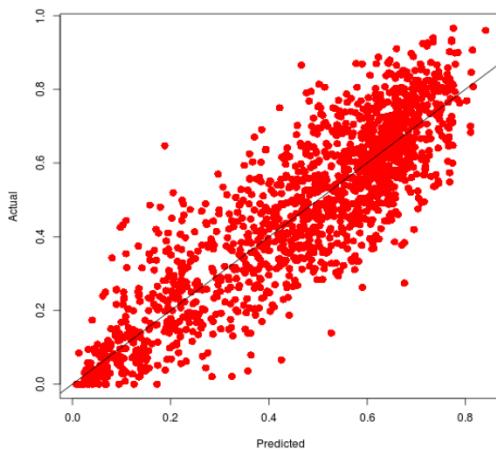


Figure 5. Validation scatterplot. For each point in the out-of-sample validation exercise, the observed value is plotted against the predicted value, with the 1:1 line shown for reference.

Model parameters	
Covariance function	Matern
Spatial scale of correlation (Km)	70.9
Variance of spatially structured component	0.5
Variance of non-spatial component	0.58
Covariate contributions	
Mean annual aridity	12%
VIIRS Nighttime Lights-2012	8%
Enhanced vegetation index	4%
SRTM Near-global Digital Elevation Models (DEMs)	14%
GPW population density	4%
Land surface temperature in the daytime	18%
Land surface temperature daily fluctuation range	4%
Land surface temperature in the nighttime	6%
Mean annual Potential Evapotranspiration	6%
Tasseled-cap brightness	6%
Tasseled-cap wetness	9%
Travel time to cities with > 50k via all transport methods	6%
Average monthly rainfall	6%
Validation statistics	
Correlation	88%
Mean absolute error	9%
Mean square error	1%

Table 2. Model fit and validation statistics. Detailed are the relative contributions of each covariate to the final model fit, and the results of the validation exercise.

### Model surface uncertainty and interpretation

The ability of the MGB process to accurately predict any given indicator is dependent on several factors. First, each indicator will have different inherent properties such as the overall amount of variation across the country, the extent to which this is spatially autocorrelated (with more autocorrelation meaning a more organized geographic pattern that is easier to predict), and the statistical distribution of values (with bi-modal, heavily skewed, or other unusual distributions often being more difficult to predict accurately). Second, the extent to which the environmental covariates are correlated with the indicator will influence predictive accuracy, with higher correlations allowing greater accuracy. Third, the density of cluster points and the sample size (e.g. number of respondents) at each cluster will have an important effect, with denser surveys and larger sample sizes yielding greater accuracy. Given these factors, it is expected that some indicators will be predicted with greater accuracy in some countries than in others. An important element of the model surface outputs is therefore the uncertainty estimates. Uncertainty takes into account the error in the mean estimate and is estimated for each grid square (pixel level). The uncertainty surface helps users understand the robustness of an estimate at any given area on the map. This uncertainty of the modeled surface is shown in Figure 2 with each grid square (pixel) value representing the width of the 95% credible interval (CI) of the mean estimate value for that pixel. It is possible to have very low confidence width values while it is also possible to have values that are equivalent to a 95% CI of 0%–100%.

#### References

Diggle, P.J., and P.J. Ribeiro. 2007. Model-based geostatistics. Edited by P. Bickel, P. Diggle, S. Fienberg, U. Gather, I. Olkin and S. Zeger. Springer series in statistics. New York: Springer.

Diggle, P.J., J.A. Tawn, and R.A. Moyeed. 1998. "Model-based geostatistics." *Journal of the Royal Statistical Society Series C-Applied Statistics* 47:299–326.

Hay, S.I., Guerra, C.A., Gething, P.W., Patil, A.P., Tatem, A.J., Noor, A.M., Kabaria, C.W., Manh, B.H.-F., Elyazar, I.A.F., Brooker, S., Smith, D.L., Moyeed, R.A., Snow, R.W. 2009. "A world malaria map: Plasmodium falciparum endemicity in 2007." *PLoS Medicine*, 6, e1000048.

Gething, P.W., A. Patil, D. Smith, C. Guerra, I. Elyazar, G. Johnston, A. Tatem, and S. Hay. 2011. "A new world malaria map: Plasmodium falciparum endemicity in 2010." *Malaria Journal* 10 (1):378.

Weiss, D.J., Atkinson, P.M., Bhatt, S., Mappin, B., Hay, S.I., Gething, P.W. 2014. "An effective approach for gap-filling continental scale remotely sensed time-series". *ISPRS Journal of Photogrammetry and Remote Sensing*, 98, 106–118.

Gething, P.W., Tatem, A.J., Bird, T., Burgett-Brucker, C.R., 2015. "Creating Spatial Interpolation Surfaces with DHS Data". DHS Spatial Analysis Reports No. 11. Rockville, Maryland, USA: ICF International.

#### Acknowledgments

Clara R. Burgett-Brucker, Samir Bhatt, Harry Gibson, Tom Pullum, Trevor Croft, Phil Satloff, Peter Getting

#### Other resources

Clara R. Burgett-Brucker, Trinadh Dontamsetti, Aileen Marshall, and Peter W. Gething. 2016. Guidance for Use of The DHS Program Modeled Map Surfaces. DHS Spatial Analysis Reports No. 14. Rockville, Maryland, USA: ICF International

### 3.3 Incertitude du modèle de surface et interprétation

Un élément important du résultat du modèle de surface est l'estimation de l'incertitude. Les cartes d'incertitude de la prédiction fournissent une indication de la précision probable de l'estimation moyenne pour chaque carré de la grille (pixel). La surface d'incertitude aide les utilisateurs à comprendre la robustesse de l'estimation à n'importe quel endroit sur la carte. L'incertitude peut varier sur une surface modélisée pour plusieurs raisons, telles que l'éparpillement des données de localisation des points de la DHS, la rareté de l'indicateur estimé et la mesure dans laquelle le modèle explique la variance. Une plus grande incertitude indique que le modèle estime mal la valeur de l'indicateur à cet endroit, tandis qu'une incertitude moindre indique que le modèle est en mesure de mieux estimer la valeur de l'indicateur à cet endroit. Cette incertitude de la surface modélisée est fournie en tant que résultat standard sous la forme d'une surface tramée (raster), chaque valeur d'un carré du quadrillage (pixel) représentant la largeur de l'IC à 95 % de la valeur de l'estimation ponctuelle de l'indicateur pour ce pixel. Il est possible d'avoir de très faibles valeurs de l'intervalle de confiance et des valeurs équivalentes à un IC à 95 % de 0 % à 100 %. Ces surfaces d'incertitude sont distinctes des statistiques de validation, qui fournissent un résumé de la performance globale du modèle plutôt que la précision à différents emplacements. Bien que la relation entre eux soit complexe, en général un modèle avec de bonnes statistiques de validation (corrélation élevée, EAM et EQM faibles) aurait également des niveaux d'incertitude plus bas, pixel par pixel et vice versa.

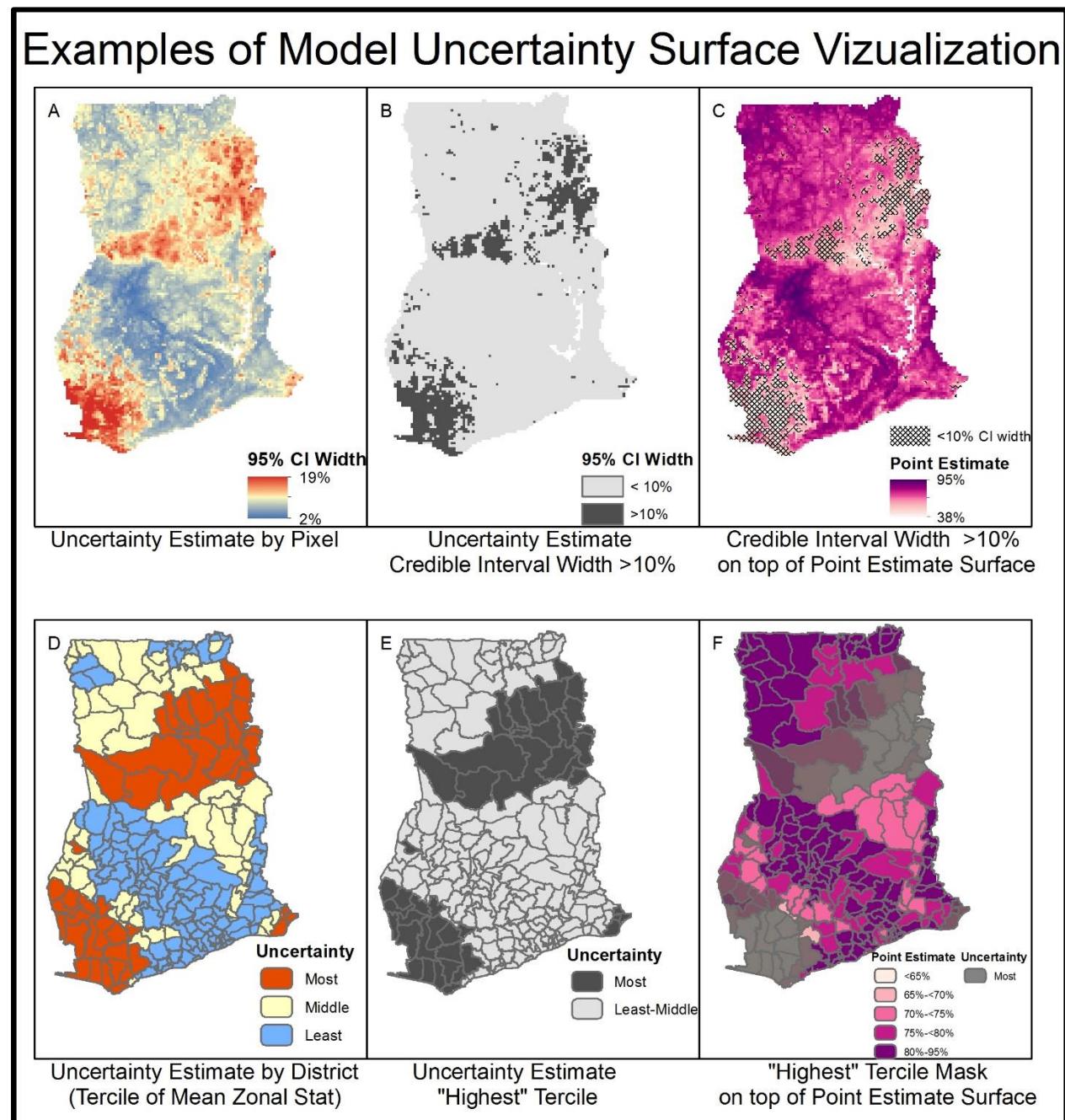
Lors de l'utilisation des surfaces modélisées pour la prise de décision, il est important de prendre en compte l'incertitude de l'estimation. L'interprétation des valeurs d'incertitude pour une surface modélisée n'est pas évidente et est souvent ignorée par les utilisateurs. Pourtant, il s'agit d'un élément important du modèle qui mérite d'être pris en considération par les utilisateurs qui peuvent en tirer une idée générale de l'incertitude du modèle et de la différence de l'incertitude entre des milieux spécifiques. Les utilisateurs doivent décider d'un niveau d'incertitude acceptable en fonction du contexte ou de l'indicateur donné et de la quantité d'incertitude qu'ils sont prêts à accepter.

La Figure 8 illustre plusieurs façons dont la surface d'incertitude peut être transformée pour être plus utile dans le processus de prise de décision. L'affichage de la surface d'incertitude sous la forme d'une valeur continue avec un schéma de couleurs divergentes (Figure 8, encadré A) permet aux utilisateurs d'identifier les zones avec plus ou moins de certitude. Une autre approche nécessite que l'utilisateur sélectionne un seuil d'intervalle crédible avec la quantité d'erreur qu'il est prêt à accepter, par exemple 10 %. La carte peut alors afficher les zones qui répondent au seuil et celles qui ne le sont pas. Il est également possible de superposer les zones qui ne respectent pas le seuil sur la carte d'estimation moyenne, soit sous la forme d'un masque ou d'une hachure (Figure 8, encadrés B et C).

Contrairement à la surface interpolée de l'estimation ponctuelle, il n'est pas approprié d'agréger (moyenne) la surface d'incertitude à des polygones géographiques plus étendus (zones) par une méthode de moyenne simple et d'utiliser directement ces résultats comme estimation de l'IC à 95 % de la valeur de l'estimation moyenne agrégée pour la même zone. Puisque l'incertitude est évaluée au niveau du pixel (la certitude de l'estimation moyenne à ce pixel seulement), elle ne peut pas être résumée par la moyenne sur une plus grande surface car le niveau d'erreur n'est pas indépendant entre les pixels. Il pourrait y avoir des façons de tenir compte de ce problème de probabilités conjointes qui se produit dans l'agrégation de la surface d'incertitude, mais il serait nécessaire de connaître l'objectif exact et le niveau d'agrégation nécessaire pour l'indicateur (Gething, Patil et Hay 2010). Ceci est approprié lors de la création d'une carte spécifique à des fins analytiques bien que le but et le niveau d'agrégation vont varier pour chaque utilisateur des surfaces modélisées du DHS Program. Il pourrait être approprié d'agréger la surface

d'incertitude à des zones géographiques plus grandes si on évalue l'incertitude relative plutôt que la largeur absolue de l'IC à 95 %. Le calcul de la moyenne de l'intervalle de confiance à 95 % pour une zone donnée et sa comparaison avec d'autres zones peut être pertinent si une zone est susceptible d'avoir plus ou moins d'incertitude par rapport aux autres zones. Opérationnaliser ceci pourrait nécessiter le calcul de la largeur moyenne de l'IC à 95 % pour les zones géographiques d'intérêt (en faisant la moyenne des valeurs de pixels dans l'unité géographique), créer des terciles de valeurs, puis les afficher comme incertitude relative élevée, moyenne et faible sans valeurs spécifiques (Figure 8, encadrés D, E et F).

Figure 8. Exemples de visualisation de surface d'incertitude du modèle





## 4 Quelles sont les limites et les hypothèses de la modélisation des surfaces ?

### Résumé

Comprendre les limites et les hypothèses de la modélisation des surfaces est essentiel pour un bon usage et une interprétation appropriée. Il y a plusieurs limites du modèle liées aux zones urbaines, à la temporalité et au biais de localisation. En outre, il peut y avoir des différences considérables dans la validité des surfaces entre différents pays et entre les indicateurs d'un même pays. Cela peut être dû à la distribution des localités échantillonées, au comptage de l'indicateur à l'échelle de la grappe et à la mesure dans laquelle les covariables sont les moteurs du processus mesuré dans ce pays. Enfin, le processus de modélisation n'a pas spécifiquement adapté le modèle pour recréer au niveau régional ou national les estimations présentes dans le rapport final de l'enquête DHS.

### Questions clés

#### *Certaines zones géographiques sont-elles modélisées mieux ou moins bien que d'autres ?*

L'activité de modélisation pilote (RAS 11) a indiqué que les zones urbaines n'étaient pas aussi bien modélisées que les autres zones d'un pays. Cela est dû en partie à la grande hétérogénéité des zones urbaines qui peut ne pas être capturée par la liste des covariables, ainsi qu'à l'impact du géomasking (déplacement) de 0 à 2 km de l'emplacement des grappes urbaines géocodées. Voir la section 4.2.

#### *Pourquoi les zones ont-elles des niveaux d'erreur différents selon les indicateurs et les pays ?*

L'erreur dans ce contexte se réfère à la capacité du modèle à prédire la valeur correcte d'un indicateur dans un endroit particulier. Cela dépend de plusieurs facteurs tels que la densité des grappes d'enquête autour d'un emplacement de prédiction, la taille de l'échantillon (nombre d'enquêtés dans chaque grappe, les points de données disponibles et leurs emplacements dans l'enquête (plus de points de données fournissent de meilleures prédictions), la force des corrélations entre les covariables et les indicateurs, l'association indicateur/covariable, et le degré inhérent de variation spatiale affiché par l'indicateur. La capacité prédictive globale du modèle est résumée dans la statistique de validation de l'erreur absolue moyenne. Voir la section 4.2.

#### *Les cartes sont-elles comparables entre différents pays ?*

Oui, les cartes sont comparables entre les différents pays mais des limites demeurent telles que décrites dans les sections précédentes; celles-ci incluent la quantité de variable d'erreurs dans la carte et la capacité prédictive du modèle. Voir la section 4.2.

#### *Puis-je recréer les estimations nationales ou infranationales du rapport final de l'EDS à partir de la carte de surface modélisée ?*

Le processus de modélisation n'a pas spécifiquement ajusté le modèle pour recréer au niveau régional ou national les estimations présentes dans le rapport final de La DHS. Dans de nombreux cas, les valeurs agrégées de la surface modélisée se situeront dans l'intervalle de confiance à 95 % des estimations du rapport final de l'EDS. Voir la section 4.3.

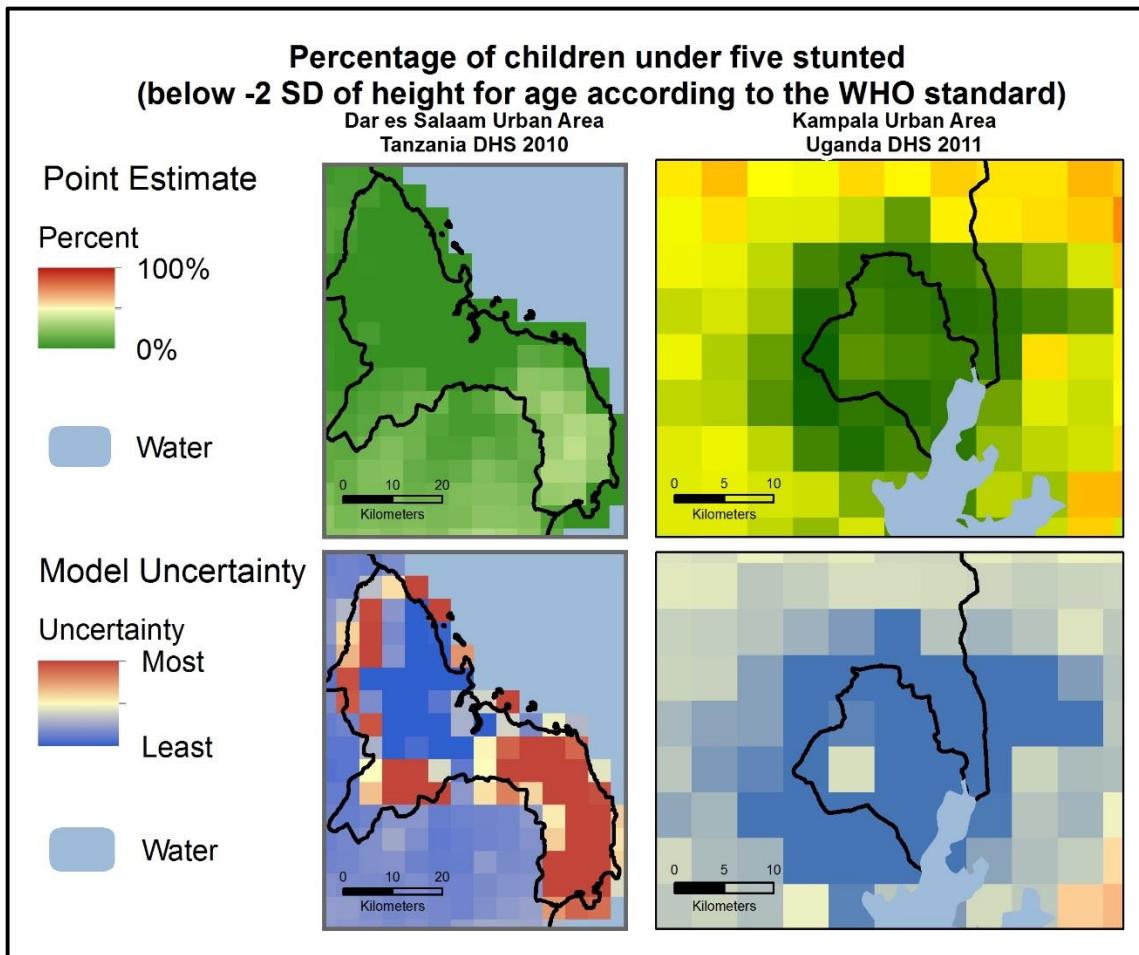
## 4.1 Les limites du modèle

Il existe plusieurs hypothèses et limites qui doivent être prises en compte lors de l'utilisation et de l'interprétation des surfaces spatialement modélisées. Celles-ci concernent les zones urbaines, la temporalité et les biais de localisation.

### 4.1.1 *Milieux urbains*

L'une des principales limites de l'approche de la modélisation est la cartographie des zones urbaines. Dans toutes les surfaces spatialement modélisées au niveau national, créées dans le projet pilote et décrites dans RAS 11, les zones urbaines ont été prédites avec des valeurs relativement uniformes (Figure 9). Cela est dû en partie à la taille de la résolution finale des pixels et à la disponibilité de covariables spécifiques au milieu urbain, qui peuvent différer de celles en milieu rural. En réalité, les grandes zones urbaines présentent généralement des hétérogénéités considérables dans les indicateurs de santé et de développement qui se produisent à des échelles plus courtes que le diamètre de pixel de 5 km. Ceci est un problème important puisque près de 50 % des résidents de la majorité des pays cartographiés habitent dans des zones urbaines et la population dans les zones urbaines est en croissance. Le rapport RAS 11 a exploré certaines approches possibles pour atténuer ces facteurs, y compris l'utilisation de covariables de résolution plus élevée dans les zones urbaines (Gething et al., 2015). Les conclusions spécifiques relatives aux zones urbaines doivent être examinées avec soin et en tenant compte du fait que la valeur prédite de  $5 \times 5$  km des pixels urbains représente une moyenne qui ne peut pas montrer la forte variabilité à l'intérieur des pixels.

**Figure 9. Exemple d'estimation ponctuelle de surface en milieu urbain et surface de l'incertitude du modèle**



#### 4.1.2 Temporalité

Il y a plusieurs limites et hypothèses temporelles ou saisonnières à prendre en compte lorsqu'on considère une surface modélisée. Ces limites sont similaires aux problèmes temporels qui existent dans le choix d'indicateurs décrit à la section 2.1.1. Les problèmes comprennent le moment de l'enquête, la durée du travail sur le terrain et la date de référence de la covariable. Ceux-ci sont particulièrement pertinents pour les indicateurs liés au temps ou ceux liés à l'époque où l'enquête a été menée, comme la saison des pluies ou l'année scolaire.

Une revue récente de 18 enquêtes EDS récentes indique que le travail sur terrain dure généralement entre 2 et 9 mois avec une moyenne de 5 mois. Les dates des enquêtes sur terrain sont disponibles sur le site Web du DHS Program (<http://dhsprogram.com/>). La durée du travail sur le terrain varie selon les pays et cette variation peut avoir un impact sur des indicateurs spécifiques d'un pays. Par exemple, si un indicateur particulier a une association saisonnière avec la date de l'enquête dans une grappe particulière, l'emplacement ne sera pas le même que dans les autres parties du pays et il se peut que des localités voisines ne soient pas enquêtées en même temps.

La période de référence des covariables est une limite supplémentaire sur les cartes de surfaces modélisées. Bien que certaines covariables géospatiales puissent ne pas varier considérablement au cours d'une année donnée ou sur plusieurs années, certaines peuvent varier considérablement. Le Tableau 2 résume la nature temporelle de chaque raster et leur (s) date (s) de collecte. L'utilisation d'une sélection de covariables standard, au lieu d'une série spécifiquement sélectionnée pour un pays donné, signifie que la date de référence d'une covariable donnée pourrait ne pas correspondre exactement à la date de l'enquête. Ceci pourrait réduire la capacité prédictive du modèle.

#### **4.1.3 Biais de localisation**

Il existe deux sources potentielles de biais de localisation dans la surface modélisée. Le premier concerne l'erreur associée à la mesure de l'emplacement du centroïde de la grappe; l'autre est l'hypothèse que l'événement mesuré par un indicateur spécifique s'est produit à l'emplacement de la grappe donnée.

Comme mentionné précédemment, les données de localisation des grappes de l'EDS utilisées dans le processus de modélisation spatiale sont un centre estimé de la grappe d'enquête, un emplacement ponctuel représentant en fait une zone de taille inconnue avec une variabilité assez importante à travers un pays, en particulier entre les zones urbaines et rurales. De plus, ces emplacements ponctuels sont géo-masqués de 0 à 2 km dans les zones urbaines et de 0 à 5 km en zones rurales, avec 1 % des emplacements ruraux déplacés jusqu'à 10 km (Burgert et al., 2013). Ces deux problèmes ajoutent une erreur spatiale au modèle bien que des travaux antérieurs décrit dans le RAS 11 aient montré que l'impact du déplacement sur l'erreur de modélisation est faible (Gething et al., 2015).

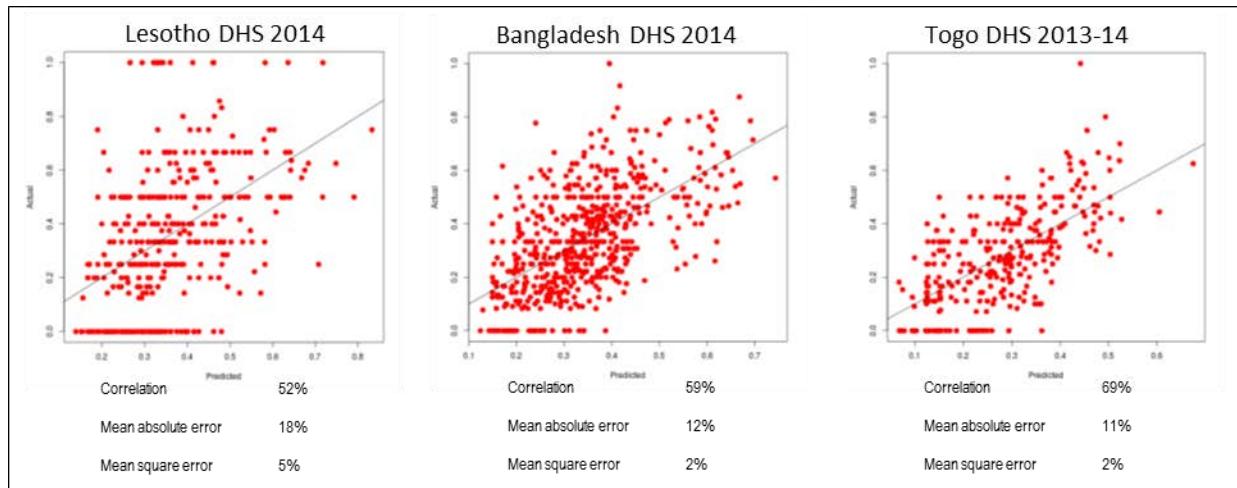
Les méthodes utilisées pour produire les surfaces modélisées supposent qu'un événement mesuré par un indicateur s'est produit à l'endroit où l'enquête a eu lieu. Cependant, l'événement aurait pu se produire dans un endroit différent. Ce type de biais se produit lorsqu'un événement se déroule dans un endroit différent de celui où a lieu l'interview. Le biais peut être minimisé si la période de référence est dûment prise en compte. Par exemple, des périodes de référence plus courtes, comme une année par rapport à une durée de vie, sont susceptibles de réduire le biais de localisation. L'objectif attendu de la surface modélisée peut également être pertinent en considérant le biais de localisation. Par exemple, si on a créé une surface interpolée de femmes âgées de 15 à 49 ans ayant subi un test de dépistage du VIH au cours des 12 derniers mois, la surface peut ou non mesurer avec précision l'impact d'une campagne de dépistage du VIH dans une zone spécifique. Certaines enquêtées ont peut-être subi le test de dépistage du VIH dans un lieu différent de celui où elles ont été interrogées. Cependant, si la surface devait être utilisée pour cibler de futures campagnes dans des zones où les tests de dépistage du VIH sont faibles, cela pourrait constituer un bon outil de ciblage. En revanche, certains indicateurs n'ont pas de biais de localisation car l'indicateur ne reflète que l'endroit où l'enquête a eu lieu. Les exemples incluent les biens des ménages, l'accès à l'eau et les pratiques d'assainissement.

### **4.2 Différences de modélisation entre indicateurs et entre pays**

La capacité du processus MBG à prédire avec précision un indicateur donné dépend de plusieurs facteurs. Premièrement, chaque indicateur a des propriétés intrinsèques différentes telles que la quantité globale de variation à travers le pays, la mesure dans laquelle celle-ci est autocorrélée spatialement (plus d'autocorrélation reflétant un modèle géographique plus organisé et plus facile à prévoir) et la distribution statistique des valeurs (avec les distributions bimodales, fortement asymétriques ou inhabituelles, plus difficiles à prévoir avec précision). Deuxièmement, la mesure dans laquelle les covariables environnementales sont corrélées avec l'indicateur influencera la précision prédictive, des corrélations plus élevées permettant une plus grande précision. Troisièmement, la densité des points de grappes et la

taille de l'échantillon (nombre de répondants) à chaque grappe auront un effet important, des relevés plus denses et des tailles d'échantillon plus grandes donnant une plus grande précision. Compte tenu de ces facteurs, certains indicateurs seront prédits avec plus de précision dans certains pays que dans d'autres. La Figure 10 montre des diagrammes de dispersion à but de validation et des statistiques de validation du retard de croissance chez les enfants de trois pays.

**Figure 10. Diagrammes de dispersion à but de validation et statistiques de validation du retard de croissance chez les enfants**



### 4.3 Agrégation de l'estimation ponctuelle de la surface interpolée au niveau national ou infra-national du DHS Program

Les surfaces modélisées peuvent s'agréger à partir de la résolution de pixel de  $5 \times 5$  km à différents niveaux administratifs ou d'autres zones géographiques. Cependant, le processus de modélisation n'a pas spécifiquement ajusté le modèle pour recréer au niveau national ou régional les estimations citées dans les rapports finaux des EDS. Dans de nombreux cas, les données agrégées devraient se situer dans l'intervalle de confiance à 95 % de l'estimation générée directement à partir des fichiers de données primaires de l'EDS. Le Tableau 5 présente un résumé des données créées dans l'étude pilote du RAS 11 sur l'indicateur de retard de croissance des enfants en Tanzanie. L'estimation issue des fichiers de données et la limite supérieure / inférieure de l'IC à 95 % représentent la valeur estimée et l'IC à 95 % obtenu lorsque l'indicateur est calculé directement à partir de l'ensemble de données recodées de l'EDS. La colonne de la valeur ponctuelle moyenne estimée est l'estimation agrégée de la surface de la carte. C'est la valeur de pixel moyenne pour cette région. Avec le retard de croissance des enfants en Tanzanie, il y a 26 régions dont 12 ayant des estimations de surface se situant dans l'IC à 95 % de l'estimation issue de l'ensemble de données (surligné en vert). Six régions étaient à moins de deux points de pourcentage de l'IC à 95 % (en jaune), alors que toutes les autres estimations se situaient entre deux points de pourcentage et dix points de pourcentage au-dessus de l'estimation de l'ensemble de données. L'évaluation des autres indicateurs en Tanzanie, au Ghana et en Ouganda, créés dans le RAS 11 a montré des résultats mixtes similaires sans tendances évidentes bien que pour certaines régions et certains indicateurs spécifiques, les estimations sont généralement meilleures que d'autres.

**Tableau 5. Estimation issue de l'ensemble des données par rapport à l'estimation issue du modèle de surface pour le retard de croissance chez les enfants en Tanzanie, par région de la DHS**

Région	Estimation issue de l'ensemble des données	Borne inférieure, ensemble des données	Borne supérieure, ensemble des données	Estimation ponctuelle de la valeur moyenne
<b>Dodoma</b>	59.0	51.7	65.9	<b>54.5</b>
<b>Arusha</b>	41.5	35.9	47.2	52.8
<b>Kilimanjaro</b>	28.9	20.3	39.3	41.8
<b>Tanga</b>	48.8	39.7	58.0	<b>47.4</b>
<b>Morogoro</b>	40.9	33.7	48.5	51.5
<b>Pwani</b>	29.3	23.0	36.5	42.9
<b>Dar es Salaam</b>	15.7	9.3	25.2	28.3
<b>Lindi</b>	50.0	41.2	58.9	<b>56.7</b>
<b>Mtwara</b>	41.2	35.1	47.5	<b>48.3</b>
<b>Ruvuma</b>	46.1	40.5	51.7	<b>51.8</b>
<b>Iringa</b>	50.8	42.1	59.4	<b>58.3</b>
<b>Mbeya</b>	41.4	29.2	54.6	<b>55.7</b>
<b>Singida</b>	38.5	32.0	45.4	<b>46.9</b>
<b>Tabora</b>	30.9	27.3	34.7	44.0
<b>Rukwa</b>	49.7	39.1	60.2	<b>51.1</b>
<b>Kigoma</b>	47.9	42.1	53.8	<b>51.8</b>
<b>Shinyanga</b>	42.5	36.9	48.3	<b>48.9</b>
<b>Kagera</b>	43.1	36.9	49.6	<b>49.2</b>
<b>Mwanza</b>	40.1	34.5	45.9	<b>41.4</b>
<b>Mara</b>	30.2	25.3	35.7	39.7
<b>Manyara</b>	46.5	40.7	52.3	<b>50.2</b>
<b>Unguja North</b>	38.8	32.2	45.8	<b>36.3</b>
<b>Unguja South</b>	25.6	19.7	32.4	<b>33.3</b>
<b>Town West</b>	19.4	14.0	26.3	<b>20.5</b>
<b>Pemba North</b>	37.3	29.3	46.0	<b>38.3</b>
<b>Pemba South</b>	29.0	23.5	35.2	38.0

## 5 Comment les surfaces modélisées du DHS Program peuvent-elles être utilisées ?

### Résumé

La section suivante donne un aperçu de l'utilisation des surfaces interpolées pour suivre et évaluer les situations et les programmes, et pour contribuer à la prise de décisions éclairées sur les politiques et les programmes futurs. Cette section inclut une discussion des approches possibles pour opérationnaliser les surfaces modélisées et les limites de ces approches.

#### Questions clés

##### *Puis-je utiliser ces cartes pour plaider en faveur du soutien aux programmes ?*

Oui, ces surfaces modélisées peuvent évaluer les zones où les programmes étaient actifs dans le passé ou identifier les zones où les programmes futurs sont nécessaires. Voir la section 5.1

##### *Puis-je comparer les zones d'intervention de mon programme à d'autres régions du pays ?*

Oui, la surface modélisée peut être résumée pour représenter des zones administratives ou d'autres zones géographiques, puis comparer aux zones de non-intervention. Voir la section 5.1.

##### *Comment puis-je rendre ceci utilisable pour mon niveau de prise de décision, le pixel de 5 × 5 km ne correspond-il pas à des zones utiles pour la prise de décision de programme ?*

Les données peuvent être agrégées à un nombre quelconque d'unités administratives de niveau supérieur, de zones d'intervention de programme ou de zones opérationnelles telles que la circonscription des établissements de santé ou les zones de subsistance. Voir la section 5.2.

##### *Les surfaces modélisées peuvent-elles être utilisées avec d'autres données telles que les centres de santé ou la densité de population pour prendre des décisions concernant un programme ?*

Oui, de nombreuses sources de données géographiques peuvent être superposées dans une même carte pour améliorer la compréhension du contexte cartographique. En outre, les données peuvent être liées spécifiquement aux points ou zones d'intervention. Voir la section 5.2.3.

## 5.1 Prise de décision avec des surfaces modélisées

Les surfaces spatialement modélisées peuvent aider de plusieurs manières à améliorer la prise de décision pour de nombreux secteurs de développement qui incluent la santé, la population, la nutrition et les programmes d'eau et d'assainissement à plusieurs niveaux.

1. Suivi et évaluation: analyse et évaluation des initiatives passées (utilisation appliquée) ou compréhension des situations existantes.
2. Planification de programme : planification future des programmes et des politiques appropriés. Certaines approches s'appliquent aux deux approches, comme l'obtention d'informations contextuelles pour améliorer la compréhension de la carte.

Les spécialistes du suivi-évaluation peuvent utiliser les données des surfaces modélisées pour évaluer les programmes passés ou bien mieux comprendre les situations existantes. De telles évaluations peuvent aider à comprendre les écarts par rapport à la norme, à attribuer une cause, ou à effectuer des évaluations d'impact qui analysent ce qui serait arrivé à la population d'une zone si un programme n'avait pas été mis en œuvre.

Le Tableau 6 résume les approches possibles pour le suivi et l'évaluation, ainsi que la planification de programmes. Celles-ci comprennent la compréhension de l'écart par rapport à la norme, la comparaison des zones d'intervention avec les zones de non-intervention, l'estimation de la charge et la liaison avec d'autres données pour une compréhension contextuelle. Les gestionnaires de programmes peuvent également utiliser les surfaces modélisées pour planifier, cibler et développer des interventions et des programmes visant à améliorer les situations dans des zones géographiques ciblées. Les interventions peuvent être ciblées plus précisément, ce qui permet d'économiser de l'argent, du temps et des ressources humaines dans la recherche des résultats les plus efficaces.

## 5.2 Transformation de la carte de surface modélisée en produits utiles (opérationnalisation)

Les surfaces modélisées peuvent être utilisées pour de nombreux types de décisions comme décrit dans la section précédente. Cependant, les surfaces doivent généralement être transformées ou opérationnalisées par l'utilisateur des données. Cette opérationnalisation peut être effectuée de plusieurs façons dont trois sont abordées dans cette section: l'agrégation, l'estimation de la charge (fardeau) et le lien avec d'autres données.

**Tableau 6. Approches de suivi et d'évaluation des programmes passés avec des surfaces modélisées**

But	Résumé	Suivi & Evaluation	Planification de Programme
<b>Comprendre les déviations par rapport à la norme</b>	Avec des problèmes sociaux et sanitaires répandus, il peut être difficile de déterminer quelles zones sont mieux loties que d'autres. En outre, il peut être difficile de discerner si les nouvelles observations à un emplacement sont anormales ou dans les limites attendues. Les mesures d'incertitude fournies dans les surfaces prédictives fournissent un moyen par lequel on peut évaluer si les différences observées entre les emplacements sont significatives dans le contexte de l'échelle plus large.	✓	
<b>Comparaison des zones d'intervention avec les zones de non-intervention (démonstration du succès du programme/obtention de défenseurs)</b>	Les évaluations d'impact comparent les résultats d'un programme à ceux d'un scénario contrefactuel pour montrer comment un domaine aurait été développé (ou aurait stagné) sans le programme; Cela valide la valeur du programme, ce qui peut aider à trouver des défenseurs pour de futures interventions.	✓	
<b>Superposer des données pour un meilleur placement contextuel</b>	Les surfaces de prévalence associées à des estimations de population à haute résolution permettent d'estimer le nombre total d'individus dans certaines catégories. Ces prédictions pourraient ensuite être utilisées avec d'autres mesures SIG et des données d'enquête telles que l'emplacement des routes ou des effectifs dans les établissements pour identifier le positionnement optimal des nouvelles écoles ou les ressources optimales nécessaires dans un nouveau centre de santé.	✓	✓
<b>Ciblage et planification de programme amélioré</b>	Les gestionnaires de programme seront en mesure de répondre à des questions telles que: Où dois-je faire des efforts pour obtenir les résultats les plus efficaces? Où se trouve le plus grand besoin pour une certaine intervention? Où sont présents certains critères pour mettre en œuvre des actions spécifiques?		✓
<b>Améliorer les estimations de la charge</b>	Actuellement, la cartographie au niveau national ou grossièrement au niveau de l'unité administrative masque des hétérogénéités et il est possible qu'elle manque des points chauds et des inégalités. Les surfaces modélisées ont le potentiel de mieux identifier ces endroits et de travailler avec des cartes de population pour quantifier et cartographier avec plus de précision la charge de morbidité et d'autres problèmes de santé.		✓

### **5.2.1 L'agrégation**

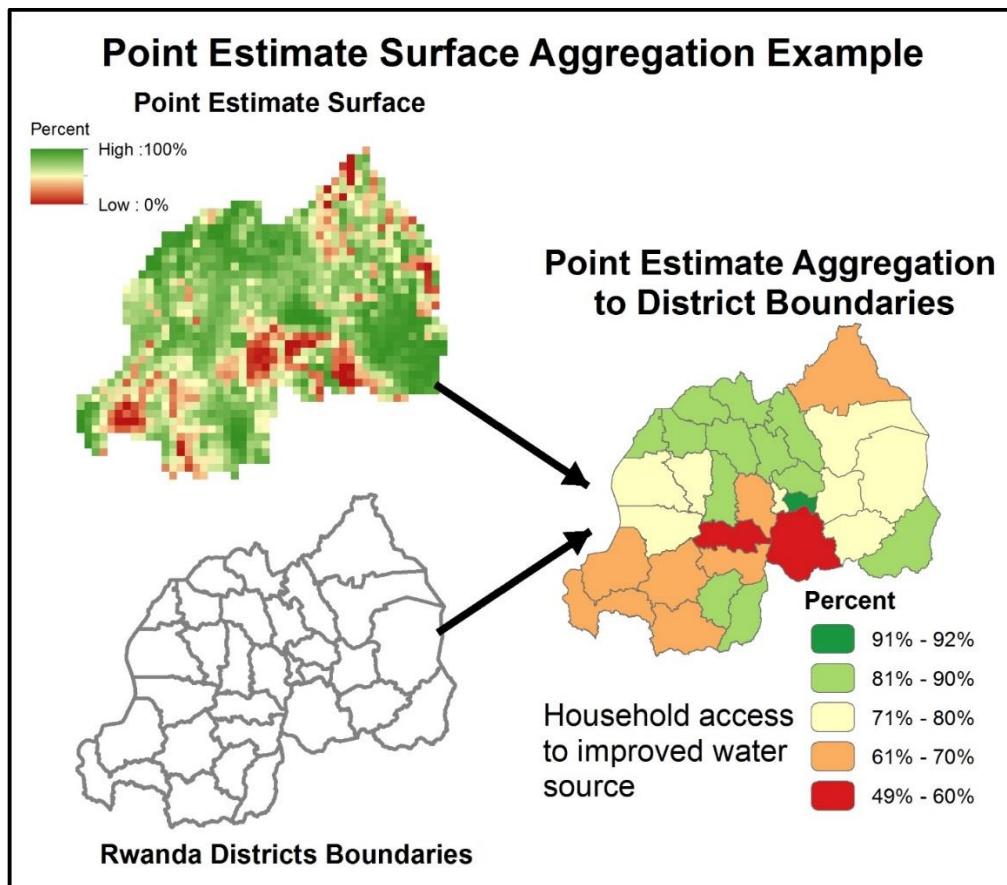
L'agrégation des estimations ponctuelles de surface modélisée, de l'échelle des pixels aux zones administratives telles que les provinces, les districts ou d'autres milieux politiquement pertinents tels que les zones de moyens de subsistance, est l'un des principaux moyens d'opérationnalisation des surfaces modélisées pour la prise de décision. L'agrégation, également appelée création de moyenne ou statistiques zonales, peut être réalisée de deux manières :

1. Statistiques zonales moyennes simples : les valeurs du polygone sont calculées en utilisant la valeur moyenne de tous les carrés de la grille (ou portions de carrés de la grille) dans la zone.
2. Statistiques zonales moyennes pondérées par la population : utilise des méthodes similaires mais prend en compte la population probable dans chaque carré de la grille et la contribution de chaque carré de la grille à l'estimation pour l'ensemble de la zone.

La Figure 11 illustre les données d'entrée du processus d'agrégation, la surface d'estimation ponctuelle et les unités administratives (dans ce cas, les districts au Rwanda). Les estimations ponctuelles sont agrégées par simple moyenne aux unités administratives pour produire une nouvelle carte qui illustre la valeur estimée moyenne pour chaque unité.

L'approche de la population pondérée peut fournir un résultat plus approprié lorsque les décisions sont axées sur l'optimisation de l'impact sur les populations. Cependant, cette approche nécessite également des données supplémentaires et des étapes de calcul. Rappelons que les données d'entrée utilisées pour la pondération par la population doivent être standardisées par rapport aux emplacements des carrés de la grille et à la taille/résolution des pixels utilisés dans les surfaces modélisées de la DHS. D'autres considérations sur les données de population sont examinées dans la section «Estimation de la charge».

Figure 11. Exemple d'agrégation de surface modélisée à l'échelle d'unités administratives



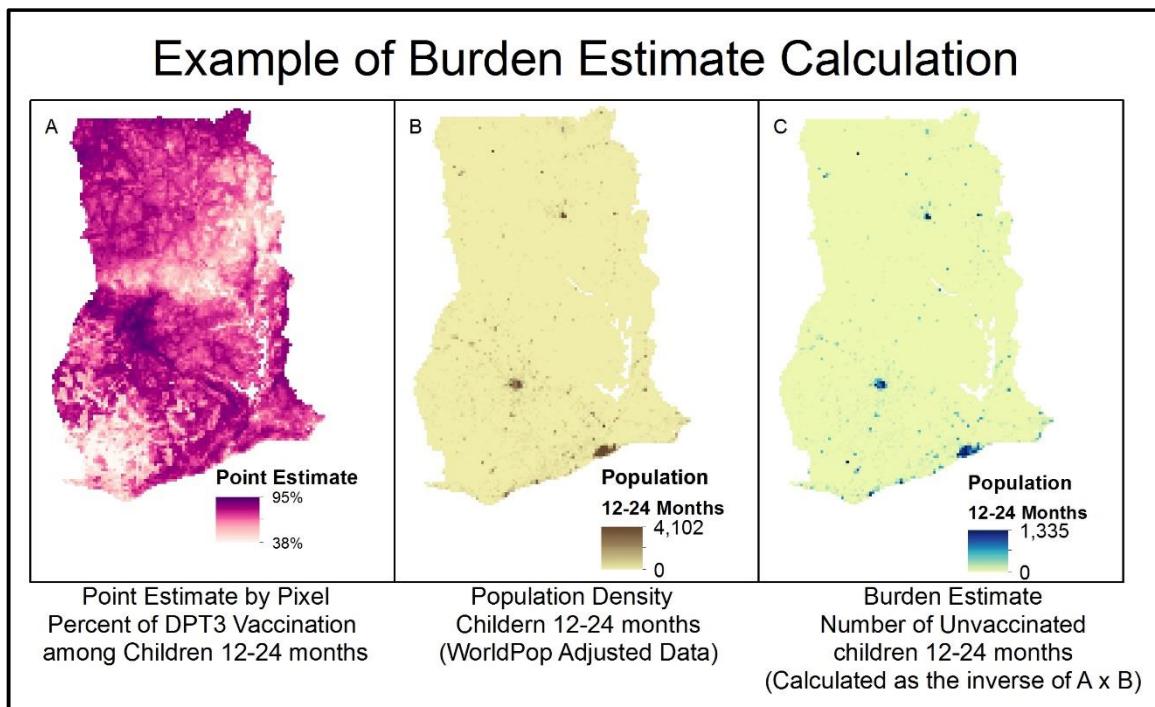
### 5.2.2 *Estimations de la charge*

Les estimations de la charge commencent avec l'estimation de la prévalence sur la surface modélisée et la conversion en nombre de population (nombre d'individus) affectée. Le choix des données de population peut influencer les résultats. Plusieurs équipes différentes ont travaillé sur la production de surfaces modélisées de la densité de la population. Il est essentiel qu'une surface de population de référence correcte soit utilisée pour l'estimation du dénominateur, ce qui inclut le pays, la tranche d'âge et le sexe. Le dénominateur doit également être limité à la population appropriée à partir de laquelle l'indicateur est estimé. Pour la vaccination, le dénominateur serait les enfants âgés de 12 à 23 mois et pour l'accouchement dans un établissement de santé, les femmes enceintes. La Figure 12 illustre le processus d'estimation de la charge pour la vaccination DTC3 au Ghana à partir des deux entrées de données (1) la surface de l'estimation ponctuelle (encadré A) et (2) la surface de la densité de population (de WorldPop) (encadré B), et le résultat final du nombre d'enfants non vaccinés (encadré C). Le nombre d'enfants non vaccinés est calculé en prenant l'inverse du produit du pourcentage d'enfants vaccinés en le multipliant par le nombre d'enfants.

Les sources de fichiers de données sur la population accessibles au public avec quelques options pour la répartition de la population et le sexe, comprennent: WorldPop (<http://www.worldpop.org.uk/>) et Gridded Population of the World (GPW) (<http://sedac.ciesin.columbia.edu/data/collection/gpw-v4>). Les données WorldPop sont des surfaces de population quadrillées qui sont ré-échantillonées et estimées en utilisant des méthodes similaires à celles utilisées dans les surfaces modélisées de la DHS qui incluent aussi

l'utilisation de covariables spatiales. Le GPW utilise une pondération réelle pour répartir la population dans l'espace avec seulement des données de recensement et sans aucune covariable supplémentaire. L'estimation de la partie de la population à risque peut être effectuée en utilisant les données de recensement (si disponibles) et les enquêtes du DHS Program ou d'autres enquêtes auprès des ménages qui incluent des données détaillées sur la structure de la population. Ces données peuvent être appliquées à la population pour estimer le nombre d'individus à risque (comme le nombre d'enfants de moins de cinq ans dans la population). Cela peut causer des problèmes de colinéarité (deux entrées du même modèle étant fortement corrélées) dans les estimations de la charge, bien que cela soit minime, étant donné le processus de modélisation global utilisé pour les surfaces de population et les surfaces modélisées du DHS Program.

**Figure 12. Illustration du processus de calcul de l'estimation de la charge**



### 5.2.3 Lien avec d'autres données

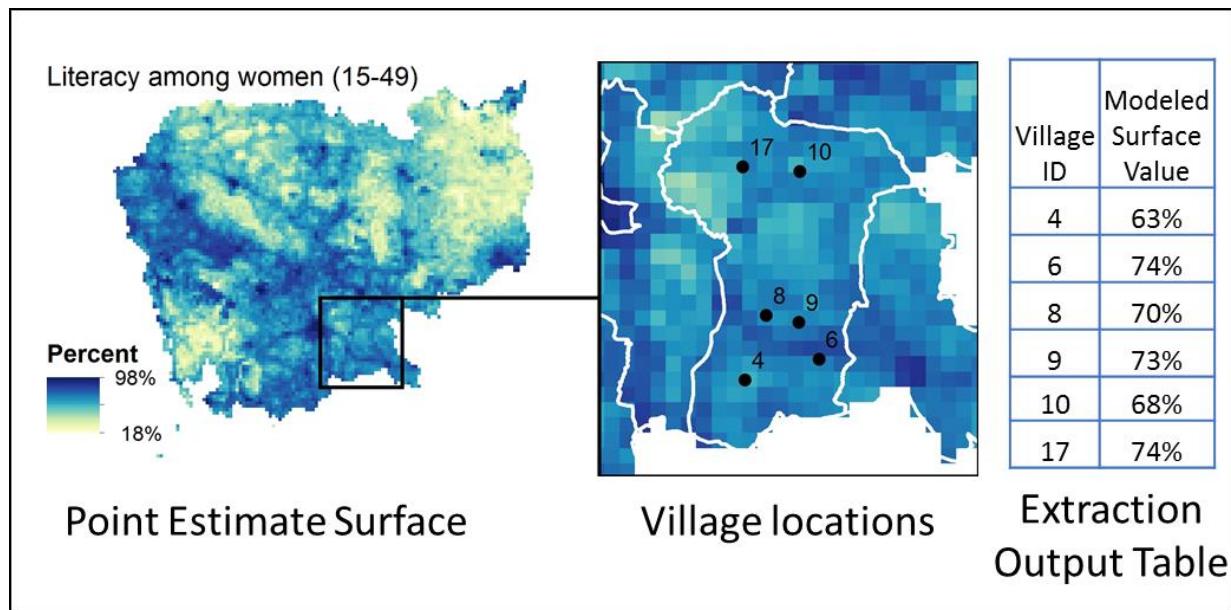
Les surfaces modélisées peuvent être liées à d'autres données géospatiales pour améliorer la compréhension du contexte et des besoins. Les deux approches les plus courantes sont l'extraction de données ponctuelles et l'inclusion des valeurs de l'estimation ponctuelle de la surface modélisée pour créer une surface modélisée analytiquement pertinente qui combine plusieurs surfaces modélisées de la DHS ou des surfaces d'autres programmes pertinents.

L'extraction de points est plus utile lorsque les données disponibles pour le couplage sont des données ponctuelles (emplacements ponctuels avec latitude et longitude connues) et lorsque les données ne proviennent pas du polygone de zone d'un programme où les techniques d'agrégation pourraient être appliquées. Par exemple, un programme a été réalisé dans certaines zones d'un district et l'emplacement GPS des villages où le programme a été mis en œuvre est disponible. Dans ce cas, une extraction de tampon peut permettre de comprendre la valeur d'un ou plusieurs indicateurs autour d'un site donné. Il est important de noter qu'une extraction ponctuelle de valeurs peut être quelque peu trompeuse, et qu'une

zone autour du point donnerait une meilleure idée du contexte. Cette valeur de tampon extraite peut ensuite être incluse comme donnée d'entrée d'une analyse plus standard.

La combinaison de plusieurs surfaces modélisées de la DHS ou la combinaison de surfaces modélisées de la DHS avec d'autres surfaces modélisées pertinentes a un potentiel analytique, quoique la colinéarité des données d'entrée doit être minitieusement prise en compte. Les surfaces modélisées du DHS Program sont créées avec une suite de covariables (voir Tableau 2 pour un résumé des covariables possibles dans le modèle, et se référer aux documentations des indicateurs pour les covariables spécifiques incluses dans chaque surface pour chaque indicateur spécifique). L'inclusion de l'une de ces covariables avec les surfaces modélisées du DHS Program pour créer une autre surface combinée pourrait conduire à une colinéarité dans les données. L'utilisation de la table des covariables dans la documentation fournira des indications sur la colinéarité qui peut être probable avec la corrélation entre la surface et une covariable similaire incluse dans le modèle.

**Figure 13. Illustration de l'extraction de la valeur de surface modélisée à la localisation des villages**



### 5.3 Considérations d'utilisation et limites

Il y a quelques considérations clés qui peuvent limiter l'utilisation des surfaces modélisées dans certaines zones ou pour certaines décisions. La section précédente a présenté quelques-unes des limites des surfaces modélisées liées à la taille des zones urbaines, à la temporalité, au biais de localisation et aux différences dans la modélisation des indicateurs et dans l'incertitude du modèle.

Les comparaisons ou les décisions concernant les grandes zones urbaines où le modèle a une homogénéité considérable ne permettent pas une compréhension à plus petite échelle des différences qui peuvent exister dans ces zones. Ainsi, une planification à échelle fine ou des décisions basées sur l'utilisation des surfaces modélisées dans ces zones serait potentiellement erronée. Cependant, la prise de décision axée sur les besoins relatifs de certaines localités par rapport à d'autres comprenant des zones urbaines pourrait être analysée comme des grandes unités combinées.

Le moment de l'enquête et la relation avec les autres données utilisées telles que les activités d'intervention du programme, doivent être soigneusement pris en compte, ainsi que la période de

référence de l'indicateur. Par exemple, si une activité de soins prénatals a débuté dans une certaine région du Ghana au début de 2013 et que le responsable du programme souhaite comparer les résultats de cette région avec d'autres régions en utilisant la surface modélisée de la DHS 2014 pour les '4 CPN ou +', les résultats seraient peu trompeurs. L'indicateur '4 CPN ou +' saisit les données sur les naissances ayant eu lieu au cours des cinq dernières années qui, dans ce cas, seraient de 1 à 2 ans du programme et de 3 années ou plus quand le programme n'était pas actif.

Lors de l'évaluation de l'impact d'un programme, il est important de savoir quels autres programmes ont pu être actifs dans les mêmes zones ou dans les zones de contrôle afin d'avoir une compréhension complète du programme au moment de l'enquête.

## Dernières réflexions

Les surfaces spatialement modélisées basées sur l'approche MBG peuvent aider à répondre aux besoins des communautés nationales et internationales pour l'obtention d'estimations plus granulaires et spatialement détaillées que celles actuellement fournies par le DHS Program et la plupart des autres sources de données au niveau national. Ces types de cartes, que ce soit à l'échelle de la grille de  $5 \times 5$  km ou agrégées par la suite aux unités décisionnelles infra-nationales appropriées, peuvent fournir des informations nécessaires pour mesurer la variation géographique des indicateurs de santé. Les surfaces spatialement modélisées du DHS Program constitueront une source d'information supplémentaire qui aidera les décideurs à mieux comprendre la désagrégation géographique des principaux indicateurs démographiques et de la santé au cours des prochaines années. Il existe un énorme potentiel pour des utilisations nouvelles et innovantes des surfaces modélisées, mais c'est seulement dans une grande communauté d'utilisateurs qui partagent leurs expériences que ce potentiel sera pleinement réalisé. Les utilisateurs sont encouragés à soumettre leurs cas et autres commentaires au DHS Program.



## Références

Burgert, C.R. 2014. *Spatial Interpolation with Demographic and Health Survey Data: Key Considerations*. DHS Spatial Analysis Reports No. 9. Rockville, Maryland, USA: ICF International. Available at <http://dhsprogram.com/pubs/pdf/SAR9/SAR9.pdf>.

Burgert, C.R., J. Colston, T. Roy, and B. Zachary. 2013. *Geographic Displacement Procedure and Georeferenced Data Release Policy for the Demographic and Health Surveys*. DHS Spatial Analysis Reports No. 7. Calverton, Maryland, USA: ICF International. Available at <http://dhsprogram.com/pubs/pdf/SAR7/SAR7.pdf>.

Diggle, P., and P.J. Ribeiro. 2007. *Model-Based Geostatistics*. Springer Science & Business Media, New York, NY.

Diggle, P.J., J. Tawn, and R. Moyeed. 1998. "Model-Based Geostatistics." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 47(3):299-350.

Gething, P., A. Tatem, T. Bird, and C.R. Burgert-Brucker. 2015. *Creating Spatial Interpolation Surfaces with Dhs Data*. DHS Spatial Analysis Reports No. 11. Rockville, Maryland, USA: ICF International. Available at <http://dhsprogram.com/pubs/pdf/SAR11/SAR11.pdf>.

Gething, P.W., A.P. Patil, and S.I. Hay. 2010. "Quantifying Aggregated Uncertainty in Plasmodium Falciparum Malaria Prevalence and Populations at Risk Via Efficient Space-Time Geostatistical Joint Simulation." *PLoS Comput Biol* (1553-7358 [Electronic]).

Gething, P.W., A.P. Patil, D.L. Smith, C.A. Guerra, I.R. Elyazar, G.L. Johnston, A.J. Tatem, and S.I. Hay. 2011. "A New World Malaria Map: Plasmodium Falciparum Endemicity in 2010." *Malar J* 10:378.

Hay, S.I., C.A. Guerra, P.W. Gething, A.P. Patil, A.J. Tatem, A.M. Noor, C.W. Kabaria, B.H. Manh, I.R. Elyazar, S. Brooker, D.L. Smith, R.A. Moyeed, and R.W. Snow. 2009. "A World Malaria Map: Plasmodium Falciparum Endemicity in 2007." *PLoS Med* 6(3):e1000048.

Weiss, D.J., P.M. Atkinson, S. Bhatt, B. Mappin, S.I. Hay, and P.W. Gething. 2014. "An Effective Approach for Gap-Filling Continental Scale Remotely Sensed Time-Series." *ISPRS Journal of Photogrammetry and Remote Sensing* 98:106-118.